

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:32:11 ; Search time 2741.47 Seconds
(without alignments)
16617.753 Million cell updates/sec

Title: US-09-828-313-22

Perfect score: 2177
Sequence: 1 accccggcgttattgctc.....gcagatatgcgttaagc 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	2177	100.0	2177	6	AX280988	AX280988 Sequence
2	540.2	24.8	683	6	AX280975	AX280975 Sequence
3	46.6	2.1	125020	2	AF429315	AF429315 Homo sapi
4	42.8	2.0	157442	2	AC099664	AC099664 Homo sapi
5	42.8	2.0	183978	2	AC024967	AC024967 Homo sapi
6	41.6	1.9	7218	6	166494	166494 Sequence 14
7	40	1.8	70203	2	AC091058	AC091058 Homo sapi
8	40	1.8	153072	2	AL360015	AL360015 Homo sapi
9	40	1.8	162508	2	AC025870	AC025870 Homo sapi
10	40	1.8	174200	2	AC022797	AC022797 Homo sapi
11	38.6	1.8	155425	30	AF429315	AF429315 Homo sapi
12	38.6	1.8	155325	30	AC013612	AC013612 Homo sapi
13	38.6	1.8	162914	2	AC107067	AC107067 Homo sapi
14	38.4	1.7	182914	2	AC008813	AC008813 Homo sapi
15	37.8	1.7	69507	2	AC009266	AC009266 Mus muscu
16	37.8	1.7	73465	2	HS256M13	AL109748 Homo sapi
17	37.8	1.7	168182	2	HS214G14	AL050303 Homo sapi
18	37.8	1.7	340000	2	HS21C004	AL163204 Homo sapi
19	37.8	1.7	6706	6	AB5311	A95311 Sequence 42
20	37.6	1.7	91940	2	AL445674	AL445674 Human DNA
21	37.6	1.7	157611	2	AC019024	AL4518024 Homo sapi
22	37.6	1.7	158307	10	AL592112	AL592112 Mouse DNA
23	37.4	1.7	179726	2	AC007052	AC007052 Homo sapi
24	37.4	1.7	181750	2	AC025506	AC025506 Homo sapi
25	37.4	1.7	192342	2	AC027588	AC027588 Homo sapi
26	37.4	1.7	196669	2	AC087535	AC087535 Homo sapi
27	37.4	1.7	200774	2	AP001592	AP001592 Homo sapi
28	37.2	1.7	155163	2	AC098990	AC098990 Rattus no
29	37.2	1.7	158825	2	AC023503	AC023503 Homo sapi
30	37.2	1.7	164481	2	AC034152	AC034152 Homo sapi
31	37.2	1.7	173794	2	AC023108	AC023108 Homo sapi
32	37	1.7	176968	9	AL161905	AL161905 Human DNA
33	36.8	1.7	138370	2	AC055707	AC055707 Homo sapi
34	36.8	1.7	152318	2	AL162851	AL162851 Homo sapi
35	36.8	1.7	169141	2	AC021538	AC021538 Homo sapi
36	36.6	1.7	171911	2	AC025637	AC025637 Homo sapi
37	36.6	1.7	62282	2	AC010304	AC010304 Homo sapi
38	36.6	1.7	62282	2	AC011347	AC011347 Homo sapi
39	36.4	1.7	165167	2	AC096938	AC096938 Rattus no
40	36	1.7	12382	2	AF207059	AF207059 Homo sapi
41	36	1.7	119000	9	AP001482	AP001482 Homo sapi
42	36	1.7	126462	2	AC004876	AC004876 Homo sapi
43	36	1.7	167491	2	AC093442	AC093442 Papio cyn
44	36	1.7	168147	9	AL139092	AL139092 Human DNA
45	36	1.7	189058	2	AC013339	AC013339 Homo sapi

ALIGNMENTS

RESULT 1	AX280988	2177 bp	DNA	linear	PAT 02-NOV-2001
LOCUS	AX280988	Sequence 22 from Patent WO0177356.			
DEFINITION	AX280988				
ACCESSION	AX280988				
VERSION	AX280988.1	GI:16608262			
KEYWORDS					
SOURCE					
ORGANISM	Physcomitrella patens.				
REFERENCE	1 (sites)				
AUTHORS	da Costa e Silva, O., Bohnert, H.J., van Thiel, N., Chen, R. and Sarrila-Millan, R.				
TITLE	Protein kinase stress-related proteins and methods of use in plants				
JOURNAL	Patent: WO 0177356-A 22 18-Oct-2001;				
FEATURES	1. 2177				
SOURCE	Location/Qualifiers				
BASE COUNT	522 a 508 c 547 g 600 t				
ORIGIN	/organism="Physcomitrella patens"				
	/db_xref="taxon:3218"				

Query Match 100.0%; Score 2177; DB 6; Length 2177;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atccccggcctctatcgctcgaataatcattgacaaatctgctgagccttcgta 60
DB 1 ATCCCGGCTGTATTGGCTCGATATTATGTAATGTAATGATTGGAGCTTCGTA 60
QY 61 ttgaagtcagcagcagcgtctgagagcttcgagcagaaagttaacacgttcgaatc 120
DB 61 TTGAGTCAGCAGCAGCAGCCTGAGACTTCGACAGCAGAACTCACTGCTGTAATTT 120
QY 121 tgaagattgaaagcgtctgaagcttcaacgttgatttggaggttcgttcgactga 180
DB 121 TGAATTTGAAGCGCGTGAAGCTTCACTGATTTGGAGGCTTTTGAAGTATGAGAG 180
QY 181 agtctctcagctgaagaaagtcttgcaatttaaggggcacaacacggttcgttcgagtcctc 240
DB 181 AGGTCCTCTGAGCTGAGAAATGTTTGCATAATTAGGGGCAACACCGGTTTGGAGTCCCTT 240
QY 241 gccacttatcaaatctgtctgattacaagctcgaagcttcaatcgaacgtlaagattt 300
DB 241 GCCACTTATTACAAATTTGTTGGTTACAGCTCGACGATTTCAATGGAACGTAGAGTTT 300
QY 301 agtcggagtcgagagatctatgfatccgctcagcagaaagaaagacccgtatctgcgaagc 360
DB 301 AGTCGGGTCGAGGATCTATGTAATCCCTCAGCGGAGAAAGAGACCTTGATCTGCCGAGC 360
QY 361 gatcgtgtgagattctgacttagaaagagttgagccgcatcagaactatctatctcttga 420
DB 361 GATGCTGTGGATTGACTAGAAAGAGGTGACCCGCACTCAACACTAATTTATTCCTTGGA 420
QY 421 gggaagagatcagagttcccaatgggtctcactcgtcttctgtgtcgaagttcaagttca 480
DB 421 GGGAGAGATCGAGGTTCCAAATGGGCTCACTCCGTTTCTTGTTGTCACGCGTTCAAGGTTA 480
QY 481 tgtccgggtgtgtctaccccccgagcagcagctcgagaatctgaagaaatctctgagctga 540
DB 481 TGTCCGGGTGTGCTACCCCGAGCGCCAGCTCGAGAAATCTGAGCAATCTTTGAGGTGCA 540
QY 541 cgaatctctctctctgggttaaaccgaactactatgtctgggttagaagacccttaacaaatcac 600
DB 541 CGATCTTCTTCTGGGTAAATCCAGACTAATCTATGTCTGCCGTAGCAACCCCTTACACANTAC 600
QY 601 caatcgtatgagcagcgaagagtgctcgagtaatgggttgaaactactcgtcttcgcaac 660
DB 601 CAATCGTATGAGCAGCGAAGAGGTGCTCGAGATGGGCTGACCTGCTTGGTTGGCGAAC 660
QY 661 gccaaatgcccacacctctctctagaaagctcagcaggaagtagtaacatcgaagga 720
DB 661 GCCAAATGCCCCAACTTCTTAGAACGTACGCCGAAGGTACTACATCGAGGATCCAAAT 720
QY 721 ttggcagaagatttccaaagatgggggtcgaatgtggggagagttcgaagccgaagcgtg 780
DB 721 TTGGCCACGATTTTCAAAACATGGGGTCCATGTGGGAGAGTGCAGAACGCCCAACCACTG 780
QY 781 gagcccaagtcagagaaagttcttgattatcaatccagtaagatgcagcagcttgaatc 840
DB 781 GAGCCCAAGTCAGAGAAAGTTTGAATATCATTTCACTAACGATGCACACCTGGAATC 840
QY 841 catcagaaacgagagcgccagagacctcaactcgtctgaagaacggaacatcgaagaacactta 900
DB 841 CATACGAACGAGAGCGCCAGAGCTCACTCCCTGGAACGACGACATTCGAACACCTTTAA 900
QY 901 gctgatttcaatctcgagcttgatctgagagcaacttgaacttccagaatttcaatagact 960
DB 901 GCTGATTTCAATCTCGGCAATTCCTTGGAGACTTTCGACTTTCAGATTTCAATAGACT 960
QY 961 aaagaagtcgcaactccctaaatctcttcggaagagccttatcgcgagcgaagcaactgcaca 1020
DB 961 AATGGAATGCGCACTCCCTAATCTTTCGAGAGGCTTATTCGCCAGCGCAACTGCGCA 1020

QY 1021 agacgagatctactcagatgatatcaaaagtcgcgaaggtccgaattaggaatgatatgt 1080
DB 1021 AGACGAGATTTACTTGAGATGATATAAAAGTCCGCAAGGTCGGAATTTGGAATGTATGT 1080
QY 1081 ttcaagaagacagagaattctatctctgaagaagcgcgtagaagcgcglaaattctgctggaa 1140
DB 1081 TTCCAAGAGACAGGAATTTCTATCTGGAAGACCGCTAGCGCGCTGAATTTGCTGGAA 1140
QY 1141 gccgggttttcgagaagatctctcgaagatgaagcctgtcaatgaattccaaactccgaatg 1200
DB 1141 GCCGGTTTTCGAGACATCTTCAGAGATGAAGCCTGTCAATGGAATTCACATCCGATGGC 1200
QY 1201 ttacgggagatgaggtctctccgggaagagcctctaccatctctacctgagccgaa 1260
DB 1201 TTACCGGAGATAGTGGGTCTCCGCGAAGAACGCTTACCCATCCTTACCTGGGCCAA 1260
QY 1261 gaaatcttaacccgcacagacagatgagtgctccgcgaagagcagtcctccgcgaagaa 1320
DB 1261 GAACATTTTCACCGCACGACAGAGTGAAGTCCCGCAAGAGAGTCCTCCCGCAAGAA 1320
QY 1321 ggtctcaacacccctcccgagccgcatcttgtagcggaactgagtgaaatctctgctgc 1380
DB 1321 GGTCTACCCACCTCCCGACGCCGATTTTGTAGCGGACTGCGTGAAGTATTGCTGCC 1380
QY 1381 atctcgaagaagttcagaagaatcgaagcagaacatctcttataatgagtgagttgtc 1440
DB 1381 ATCTCAGACAGTTTCAACCAATCGAGCAACCGCAAACTCTTTATATGCGCTAGTTGT 1440
QY 1441 gttctgagctgaactctctatctatctcccatcagagataactcgtctgataaattc 1500
DB 1441 GTCTGCACTGAAGCTCTTATCTATTCCTCCATCGAGATACGATTTGTTGATTAATTT 1500
QY 1501 ctccaacattctgtctctcctaccccaagagctctccaatgagcgaataatgtaacga 1560
DB 1501 CTCCAACTTTTGTGCTTTCATCTCAAGACAGCTCTCAAGGCGCAATATGTGTAGGA 1560
QY 1561 catgtgcacaactccaaatcagtagagtgatctctgtaaccacagttcactgaagttatca 1620
DB 1561 CATGTGCACAACCTCAATTAAGTAGAGTTATCTGTAAACCAAGTTCTGAGGTATCA 1620
QY 1621 aggaatggcgagtagaagactgacttctgtcttgatcccggtctgtgaagaatgtc 1680
DB 1621 AGGAATGCCAGTAAGCACTGCTACTTGTGCTTGTGTATCCGTTGTGAGAAATGTC 1680
QY 1681 atgtgcagcgtgctctacagtgagatctctctgaagcgagatctctgcttcgcaagttg 1740
DB 1681 ATGTGCAGCCTGCTTATCACTGAGGATTTCTTGAAGCCAGATCTTGTCCGAGTTTG 1740
QY 1741 ttctataagctttctgtctglaaggggcttaagcgtlactcaagcaatgaagaagttg 1800
DB 1741 TTCTATPACGTTTGTGTTGTTAGGGGCTTAACAGCTACTATCAACCAATGAGAGTGTG 1800
QY 1801 ctggtgtgagatttgaacgaactctcttggagagatgtctcttccatgtatgaacaaga 1860
DB 1801 CTGGTGTGATTTGACACAAATCTTTTGGAGATTTGCTTTCCATATGTGAACAATAGGA 1860
QY 1861 ggaacactgagcctgtgtggtgcacatcccatagaacaatgacttcaacttcgggtgtctc 1920
DB 1861 GGACACTGTGGCTGTGTGGGCACTCCATAGAACATAGTGTCTTCACTTGTGGGTGTC 1920
QY 1921 accactagaatcataagactctctactatcttccgggtctgtcttcgaagctcagtlacc 1980
DB 1921 ACCACTAGATATATGACCTTCTCATATTAATTTGGGCTTGTTCGAGCTCATGTATCC 1980
QY 1981 atcgactagaagctcaacttgaactgagtgataatctgttgcaaatctgagtgaggtcttga 2040
DB 1981 ATGCACTAGAGCTCACTTTGACTGAGTGGATATCGTTTGTCATTTAGTGGAGCTTGTGA 2040
QY 2041 gatgatatagtgcaattgtacagtagctgtgaatgtctgtttacaagaatgagcagctaga 2100
DB 2041 GATGATAGATGCCATTTGTATCACTAGTGTGATGCTGTTTACAAAGATAGCGGCACTAGA 2100

[illegible]

RESULT / LOCUS	AC099664/C
DEFINITION	AC099664 Homo sapiens chromosome 3 clone RP11-6B4, WORKING DRAFT SEQUENCE, 2
ACCESSION VERSION	AC099664 AC012585 AC099664.1 GI:16973745
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 157442) Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D. Direct Submission Unpublished 2 (bases 1 to 157442) Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D. Direct Submission Direct Submission Genome Center, University of Washington, Box 352115, Seattle, WA 98195, USA On Nov 17, 2001 this sequence version replaced gi:13194374.
JOURNAL	
TITLE	
AUTHORS	
REFERENCE	
COMMENT	

```

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: BCM

----- Project Information -----
Center project name: chr-3
Center clone name: RP11-6B4 (bc0602)

----- Summary Statistics -----
Sequencing vector: unknown: 44% of reads
Sequencing vector: plasmid: 108752: 56% of reads
Chemistry: Dye-terminator: 89% of reads
Chemistry: Dye-terminator: Big Dye: 11% of reads
Assembly program: Phrap, version 0.990119
Consensus quality: 155430 bases at least Q40
Consensus quality: 155673 bases at least Q30
Consensus quality: 157102 bases at least Q20
Insert size: 176750: 14.0% error: agarose-tp
Insert size: 157342: sum-of-confliqs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 65404: contig of 65404 bp in length
* 65405 65504: gap of unknown length
* 65505 157442: Contig of 91938 bp in length.

FEATURES
    source
        1..157442 "Homo sapiens"
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-6B4"
            /clone_lib="RPCI human BAC library 11"
            1..65404
            /note="assembly_name:Contig35"
            65505..157442
            /note="assembly_name:Contig36"

BASE COUNT      50491 a 28472 c 27460 g 50919 t      100 others
ORIGIN

```

Query Match	2.0%	Score 42.8	DB 2	Length 157442
Best Local Similarity	48.4%	Pred. No. 0.48		
Matches 119	Conservative 0	Mismatches 127	Indels 0	Gaps 0

QY	1891	agaacactgctcttcactcttggtgtgtgtcaccaactagatcatatgaccttcaatca	1550
DB	49206	AGAACTACTACTGCAATTTGTATTTTCTCTCGTATATTTTGTATGTCCTTATTCCTC	49147
QY	1951	ttctcgggctctgtgttcgagatcatgtacacatgcacagcgcacttgaacgcggtat	2010
DB	49146	TCTTGCTGCTCTCCCTTGAGATTAATTTTGTGTCTAGTAGTCTCTATGATTTTACTTTT	49087
QY	2011	aatcgtctgttcaattctgtgtgagcttctgtatgataagatgccaattgtacagtagctgt	2070
DB	49086	TATCTTTTGTGCATATACTATTAAGTTTTTGTCTGTGTACCATATGAGGCTTACTATAAA	49027
QY	2071	gatgcgtgttcaacaagaatgcgcagctagaagaccttcaacctcttagctaccatattat	2130
DB	49026	CATGCTATATCAAAAACAGGGGCTTTACACCATATACCACTTAACTTTGATTTGCATAAA	48967
QY	2131	tttaac	2136
DB	48966	ATTATTC	48961

RESULT	5	
AC024967		
LOCUS		183978 bp DNA linear HTG 07-MAY-2001
DEFINITION	Homo sapiens chromosome 3 clone RP11-554B18 map 3,	WORKING DRAFT
ACCESSION	AC024967	
VERSION	AC024967.4	GI:13959248
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human;	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Euarcharia; Primates; Carnivora; Homnidae; Homo.	
TITLE	1 (bases 1 to 183978)	
JOURNAL	Britten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,	
REFERENCE	Homo sapiens chromosome 3, clone RP11-554B18	
AUTHORS	Unpublished	
	2 (bases 1 to 183978)	
	Britten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,	
	Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,	

6817	7597:	conting of 781 bp	in length
7598	7697:	gap of 100 bp	
7698	8448:	conting of 752 bp	in length
8450	8549:	gap of 100 bp	
8550	9327:	conting of 763 bp	in length
9313	9412:	gap of 100 bp	
9413	10274:	conting of 762 bp	in length
10175	10274:	gap of 100 bp	
10275	11033:	conting of 759 bp	in length
11033	11133:	gap of 100 bp	
11133	11910:	conting of 777 bp	in length
11911	12010:	gap of 100 bp	
12011	12722:	conting of 712 bp	in length
12723	13529:	conting of 707 bp	in length
13530	13629:	gap of 100 bp	
13630	14358:	conting of 729 bp	in length
14359	14458:	gap of 100 bp	
14459	15238:	conting of 780 bp	in length
15233	15338:	gap of 100 bp	
15339	16049:	conting of 711 bp	in length
16050	16149:	gap of 100 bp	
16150	16903:	conting of 754 bp	in length
16904	17009:	gap of 100 bp	
17009	17752:	conting of 749 bp	in length
17753	17852:	gap of 100 bp	
17853	18576:	conting of 724 bp	in length
18577	18676:	gap of 100 bp	
18677	19399:	conting of 719 bp	in length
19399	19495:	gap of 100 bp	
19496	20253:	conting of 758 bp	in length
20254	20353:	gap of 100 bp	
20354	21099:	conting of 746 bp	in length
21100	21199:	gap of 100 bp	
21200	21917:	conting of 718 bp	in length
21918	22017:	gap of 100 bp	
22018	22687:	conting of 670 bp	in length
22688	22787:	gap of 100 bp	
22788	23499:	conting of 711 bp	in length
23499	23598:	gap of 100 bp	
23599	242318:	conting of 720 bp	in length
242319	24416:	gap of 100 bp	
24419	252180:	conting of 762 bp	in length
252181	25280:	gap of 100 bp	
25281	26009:	conting of 729 bp	in length
26010	26109:	gap of 100 bp	
26110	26865:	conting of 756 bp	in length
26866	26965:	gap of 100 bp	
26966	27745:	conting of 780 bp	in length
27746	27845:	gap of 100 bp	
27846	28555:	conting of 711 bp	in length
28557	28655:	gap of 100 bp	
28657	29411:	conting of 758 bp	in length
29415	29511:	gap of 100 bp	
29515	30240:	conting of 726 bp	in length
30241	30340:	gap of 100 bp	
30341	31109:	conting of 769 bp	in length
31111	31209:	gap of 100 bp	
31210	31929:	conting of 720 bp	in length
31930	32029:	gap of 100 bp	
32030	33755:	conting of 723 bp	in length
33755	33852:	gap of 100 bp	
33853	33553:	conting of 701 bp	in length
33554	33653:	gap of 100 bp	
33654	34362:	conting of 709 bp	in length
34363	34462:	gap of 100 bp	
34463	35228:	conting of 764 bp	in length
35227	35328:	gap of 100 bp	
35327	36088:	conting of 762 bp	in length
36088	36188:	gap of 100 bp	
36189	35927:	conting of 739 bp	in length
35928	37027:	gap of 100 bp	
37028	37761:	conting of 734 bp	in length

37762	37861:	gap of	100	bp
37862	38616:	contig of 755	bp	In length
38617	38716:	gap of	100	bp
38717	393475:	contig of 759	bp	In length
39476	39575:	gap of	100	bp
39576	40322:	contig of 754	bp	In length
40330	40423:	gap of	100	bp
40430	41154:	contig of 725	bp	In length
41155	41254:	gap of	100	bp
41255	41987:	contig of 733	bp	In length
41988	42087:	gap of	100	bp
42088	42805:	contig of 718	bp	In length
42806	42905:	gap of	100	bp
42906	43656:	contig of 751	bp	In length
43657	43756:	gap of	100	bp
43757	44475:	contig of 719	bp	In length
44476	44575:	gap of	100	bp
44576	45313:	contig of 738	bp	In length
45314	45413:	gap of	100	bp
45414	46145:	contig of 732	bp	In length
46146	46245:	gap of	100	bp
46246	46996:	contig of 751	bp	In length
46997	47096:	gap of	100	bp
47097	47683:	contig of 737	bp	In length
47684	47933:	gap of	100	bp
47934	48685:	contig of 722	bp	In length
48686	48735:	gap of	100	bp
48736	49474:	contig of 719	bp	In length
49475	49574:	gap of	100	bp
49575	50310:	contig of 736	bp	In length
50311	50410:	gap of	100	bp
50411	51135:	contig of 725	bp	In length
51136	51235:	gap of	100	bp
51236	51944:	contig of 709	bp	In length
51945	52044:	gap of	100	bp
52045	52755:	contig of 711	bp	In length
52756	52855:	gap of	100	bp
52856	53627:	contig of 772	bp	In length
53628	53727:	gap of	100	bp
53728	54482:	contig of 755	bp	In length
54483	54582:	gap of	100	bp
54583	55525:	contig of 743	bp	In length
55526	55425:	gap of	100	bp
55426	56168:	contig of 743	bp	In length
56169	56268:	gap of	100	bp
56269	56981:	contig of 713	bp	In length
56982	57081:	gap of	100	bp
57082	57844:	contig of 763	bp	In length
57845	57944:	gap of	100	bp

```
Query Match      1.88; Score 40; DB 2; Length 70203
```

Best Local Similarity 58.3%; Pred. No. 3.1;

Matches	70;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

[illegible]

RESULT	8
AL360015/C	
LOCUS	133072 bp DNA linear PRI 23-JAN-2001
DEFINITION	Human DNA sequence from clone RP11-344F20 on chromosome 6, complete
ACCESSION	AL360015
VERSION	AL360015.25 GI:12539717
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
Mammalia; Eutheria; Primates; Carniata; Vertebrata; Euteleostomi; 1 (bases 1 to 153072) Mail,M. Direct Submission Submitted (22-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquyer@sanger.ac.uk On Jan 26, 2001 this sequence version replaced g112227402. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chrf6 RP11-344F20 is from the library RP11-1.2 constructed by the group of Pleter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBAC3.6 IMPORTANT: This sequence is not the entire insert of clone RP11-344F20. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-344F20 is at 153072 in this sequence. The true right end of clone RP11-505K1 is at 100 in this sequence.	
FEATURES source	Location/Qualifiers 1..153072 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /clone="RP11-344F20" /clone_id="RP11-11.2" 60..1177 /note="MIR repeat: matches 28..144 of consensus" 714..875 /note="MIR repeat: matches 80..252 of consensus" 1068..1143 /note="L2 repeat: matches 2641..2717 of consensus" 3293..3530 /note="MER8 repeat: matches 1..239 of consensus" 4107..4180 /note="37 copies 2 mer ac 91% conserved" 4421..4732 /note="AluY repeat: matches 1..310 of consensus" 4736..5207 /note="MIR1H repeat: matches 33..541 of consensus" 6051..6176 /note="63 copies 2 mer aa 59% conserved" 6629..7114 /note="L2 repeat: matches 2238..2746 of consensus" 7120..7215 /note="48 copies 2 mer ta 69% conserved" 8053..8404 /note="MIR1A1 repeat: matches 12..365 of consensus" 8872..9833 /note="L1P4S repeat: matches 5184..6143 of consensus" 11814..14211 /note="L1P47 repeat: matches 3692..6120 of consensus" 15010..15150
	 /note="MIR repeat: matches 73..218 of consensus" 15163..15543 /note="MST4 repeat: matches 1..426 of consensus" 15800..15900 /note="MER33 repeat: matches 33..135 of consensus" 15903..16040 /note="L2 repeat: matches 2183..2324 of consensus" 16259..16488 /note="L2 repeat: matches 1818..2033 of consensus" 16560..16623 /note="MER96 repeat: matches 1..60 of consensus" 16918..17030 /note="MER91C repeat: matches 1..119 of consensus" 17234..17285 /note="26 copies 2 mer ag 88% conserved" 17305..17526 /note="MIR repeat: matches 22..258 of consensus" 18236..18606 /note="L1MD2 repeat: matches 5494..5968 of consensus" 18607..18978 /note="L1MD2 repeat: matches 5951..6321 of consensus" 20104..20471 /note="L1P4I2 repeat: matches 4063..4405 of consensus" 20472..20828 /note="THE1C repeat: matches 3..371 of consensus" 20830..22459 /note="THE1C-internal repeat: matches 4..1578 of consensus" 22463..22827 /note="THE1C repeat: matches 1..371 of consensus" 22828..24496 /note="L1P4I2 repeat: matches 4405..6153 of consensus" 24508..24802 /note="AluY repeat: matches 1..285 of consensus" 25079..25177 /note="U6 repeat: matches 1..99 of consensus" 25187..25625 /note="L1P repeat: matches 2128..2598 of consensus" 25614..25970 /note="L1P4I5 repeat: matches 5128..5472 of consensus" 25971..26271 /note="AluY repeat: matches 1..300 of consensus" 26272..26917 /note="L1P4I5 repeat: matches 5472..6153 of consensus" 27112..27147 /note="18 copies 2 mer tt 88% conserved" 27433..27482 /note="25 copies 2 mer tt 76% conserved" 27489..28511 /note="L1M48 repeat: matches 5254..6288 of consensus" 28518..28859 /note="THE1B repeat: matches 3..364 of consensus" 28860..30410 /note="THE1B-INTERNAL repeat: matches 1..1580 of consensus" 30416..30731 /note="THE1B repeat: matches 15..364 of consensus" 30733..31170 /note="L1M48 repeat: matches 4823..5253 of consensus" 31622..31890 /note="L1M6C repeat: matches 1738..1988 of consensus" 31891..32097 /note="MER6B repeat: matches 2..208 of consensus" 32098..32360 /note="L1M6C repeat: matches 1988..2256 of consensus" 33222..33283 /note="MER96 repeat: matches 115..172 of consensus" 34160..34490 /note="MER7A repeat: matches 1..346 of consensus" 38388..38597 /note="MIR repeat: matches 31..250 of consensus" 41237..41312 /note="MIR repeat: matches 49..124 of consensus"


```

repeat_region 41354..41649 /note="L2 repeat: matches 1391..1699 of consensus"
repeat_region 41651..41803 /note="L2 repeat: matches 1432..1586 of consensus"
repeat_region 41814..42061 /note="L1PA16 repeat: matches 5895..6157 of consensus"
repeat_region 42251..42816 /note="L1PA15 repeat: matches 5547..6153 of consensus"
repeat_region 42824..43089 /note="L2 repeat: matches 1573..1855 of consensus"
repeat_region 43151..43249 /note="L2 repeat: matches 2431..2517 of consensus"
repeat_region 43272..43826 /note="L2 repeat: matches 1885..2461 of consensus"
repeat_region 43827..44123 /note="AluY repeat: matches 12..309 of consensus"
repeat_region 44124..44159 /note="L2 repeat: matches 2461..2494 of consensus"
repeat_region 44633..44731 /note="L2 repeat: matches 2430..2516 of consensus"
repeat_region 45231..45525 /note="L2 repeat: matches 1..295 of consensus"
repeat_region 46431..46555 /note="L1PA3 repeat: matches 6027..6146 of consensus"
repeat_region 46742..46959 /note="L1MA3A repeat: matches 6071..6291 of consensus"
repeat_region 47659..47764 /note="L2 repeat: matches 2619..2727 of consensus"
repeat_region 49164..49564 /note="L2 repeat: matches 5762..6184 of consensus"
repeat_region 52294..52539 /note="L1MB3 repeat: matches 5762..6184 of consensus"
repeat_region 53737..54225 /note="AluX repeat: matches 1..306 of consensus"
repeat_region 55380..55708 /note="L2 repeat: matches 2223..2743 of consensus"
repeat_region 55720..55804 /note="MER33 repeat: matches 2..324 of consensus"
repeat_region 57070..57366 /note="MER5A repeat: matches 4..101 of consensus"
repeat_region 59201..59494 /note="AluY repeat: matches 1..297 of consensus"
repeat_region 59840..59893 /note="AluY repeat: matches 6..300 of consensus"
repeat_region 61297..61474 /note="28 copies 2 mer to 76% conserved"
repeat_region /note="MER5B repeat: matches 1..178 of consensus"

Query Match 1.8%; Score 40; DB 9; Length 153072;
Best Local Similarity 58.3%; Pred. No. 3.5;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1897 tagtgcttcactcgggtgttcaccacatgatatgacctctcatatttcg 1956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103343 TACGCTTCATTTAGCTTCCTTCACAAATATGACATGACCTCTCTTTTAA 103284

QY 1957 ggccttgcttcgagctcatgacatgcactagcgtcacttcgtggtgataatcgt 2016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103283 ATTAATGCTTTTAGCTCAATATCTGTGACAGATCACTTCGACGCTTGAGTAATTGT 103224

```

```

RESULT 9
AC025870 162508 bp DNA linear HTG 20-SEP-2000
LOCUS Homo sapiens clone RP11-25P4, WORKING DRAFT SEQUENCE, 42 unordered
DEFINITION
pieces.
AC025870
VERSION AC025870.2 GI:10198533
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162508)

```

AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Boguslavsky, L., Bouhagalter, B., Brown, A., Burkett, G.,
 Campilongo, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.,
 Dodge, S., Domino, M., Doyle, M., Ferrel, P., Filtrugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karakas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Lettre, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGuire, A., McKenna, K., Miengo, V., Morrow, J.,
 Meldrum, J., Menes, L., Milnova, T., Miranda, C., O'Connor, T., O'Donnell, P.,
 Murphy, T., Naylor, J., Norman, C., H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olliver, T., M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vasiliy, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 20, 2000 this sequence version replaced gl:7249393.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L7870
 Center clone name: 25_P4

Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960721
 Consensus quality: 14193 bases at least Q40
 Consensus quality: 152022 bases at least Q30
 Consensus quality: 156043 bases at least Q20
 Insert size: 151000; agarose-fp
 Insert size: 158408; sum-of-ctrls
 Quality coverage: 3.3 in Q20 bases; sum-of-ctrls
 Quality coverage: 3.2 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
 consists of 42 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1133: contig of 1133 bp in length
 1134 1233: gap of 100 bp
 1234 2439: contig of 1206 bp in length
 2440 2539: gap of 100 bp
 2540 4058: contig of 1519 bp in length
 4059 4158: gap of 100 bp
 4159 5400: contig of 1242 bp in length
 5401 5500: gap of 100 bp
 5501 18786: contig of 13286 bp in length
 18787 18886: gap of 100 bp
 18887 18943: contig of 1057 bp in length
 18944 20043: gap of 100 bp
 20044 21163: contig of 1120 bp in length

21164	21263:	gap of	100	bp	in length
21264	22264:	contig of 1391	bp	in length	
22655	22754:	gap of	100	bp	
22755	24230:	contig of 1536	bp	in length	
24231	24300:	contig of	100	bp	
24391	25594:	contig of 1204	bp	in length	
25391	25694:	gap of	100	bp	
25695	27397:	contig of 1703	bp	in length	
27398	27497:	gap of	100	bp	
27498	29330:	contig of 1833	bp	in length	
29331	29430:	contig of	100	bp	
29431	31357:	contig of 1927	bp	in length	
31358	31457:	gap of	100	bp	
31458	33360:	contig of 1903	bp	in length	
33361	33460:	gap of	100	bp	
33461	36004:	contig of 2604	bp	in length	
36005	36164:	gap of	100	bp	
36165	37979:	contig of 1815	bp	in length	
37980	38079:	gap of	100	bp	
38080	40463:	contig of 2384	bp	in length	
40464	40563:	gap of	100	bp	
40564	43574:	contig of 3011	bp	in length	
43575	43674:	gap of	100	bp	
43675	45259:	contig of 1585	bp	in length	
45260	45359:	gap of	100	bp	
45360	47385:	contig of 2026	bp	in length	
47386	47485:	gap of	100	bp	
47486	49679:	contig of 2194	bp	in length	
49680	49797:	gap of	100	bp	
49789	51819:	contig of 2040	bp	in length	
51820	51919:	gap of	100	bp	
51920	55270:	contig of 3351	bp	in length	
55271	55370:	gap of	100	bp	
55371	58223:	contig of 2853	bp	in length	
58224	58323:	gap of	100	bp	
58324	61559:	contig of 3246	bp	in length	
61570	61669:	gap of	100	bp	
61670	65776:	contig of 4107	bp	in length	
65777	65876:	gap of	100	bp	
65877	68164:	contig of 2288	bp	in length	
68165	68264:	gap of	100	bp	
68265	72330:	contig of 4066	bp	in length	
72331	72430:	gap of	100	bp	
72431	75721:	contig of 3291	bp	in length	
75722	75821:	gap of	100	bp	
75822	79327:	contig of 3506	bp	in length	
79328	79427:	gap of	100	bp	
79428	86170:	contig of 6743	bp	in length	
86171	86270:	gap of	100	bp	
86271	92478:	contig of 6208	bp	in length	
92479	92578:	gap of	100	bp	
92579	98667:	contig of 6089	bp	in length	
98668	98767:	gap of	100	bp	
98669	105029:	contig of 6262	bp	in length	
105030	105129:	gap of	100	bp	
105130	111321:	contig of 6192	bp	in length	
111322	111421:	gap of	100	bp	
111422	117477:	contig of 6056	bp	in length	
117478	117577:	gap of	100	bp	
117578	125755:	contig of 8178	bp	in length	
125756	125855:	gap of	100	bp	
128856	133343:	contig of 7488	bp	in length	
133344	133443:	gap of	100	bp	
133444	141679:	contig of 8236	bp	in length	
141680	141769:	gap of	100	bp	
141769	153366:	contig of 11867	bp	in length	
153667	153766:	gap of	100	bp	
153767	162034:	contig of 8268	bp	in length	
162035	162134:	gap of	100	bp	
162135	162508:	contig of 374	bp	in length	

FEATURES
source

	Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
	1.8%; Score 40; DB 2; Length 162508;						
	58.3%; Pred. No. 3.6;						
	70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;						
Ox	1897	tagctcttcacattcggctgtgtcgtccacagctgagtcacatgatgaccttcacatcttcttg	1956				
Db	88829	TATCTCTTCATTTAGACCTTCTTTCACACATATATGAAGAAGACCTCTCTCTTTTAA	88888				
Ox	1957	ggctctgttcagagctcatgtaccatgacctagcgcgtcacttgaactgcgtgataatcgt	2016				
Db	88889	ATTAAATGTTTTTACCTCAATTTATCTGTGACAAAGATCACTTCGCAGCTTGCGGAATTGT	88948				
RESULT 10							
AC022797							
LOCUS	AC022797	174200 bp	DNA	linear	HTG 16-MAR-2000		
DEFINITION	Homo sapiens chromosome 6 clone RP11-281G13 map 6, WORKING DRAFT						
ACCESSION	AC022797	22 unoriented pieces.					
VERSION	AC022797.3	GI:7249244					
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
TITLE	1 (bases 1 to 174200)						
COMMENT	Biren,B., Linton,L., Nusbaum,C. and Lander,E.						
JOURNAL	Homo sapiens chromosome 6, clone RP11-281G13						
UNPUBLISHED							

REFERENCE
AUTHORS

2 (bases 1 to 174200)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F., Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Geroy, S., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J., Landers, J., Lechoczky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Telamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced g1:6984458.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Genome Center

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L6043

Center clone name: 281.G.13

Sequencing vector: M13: M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163082 bases at least Q40

Consensus quality: 168787 bases at least Q30

Consensus quality: 170797 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 172100; sum-of-coverage

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently

consists of 22 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 3194: contig of 3194 bp in length

3195 3294: gap of 100 bp

3295 5794: contig of 2500 bp in length

5795 5894: gap of 100 bp

5895 9861: contig of 3967 bp in length

9862 9961: gap of 100 bp

9962 13464: contig of 3503 bp in length

13465 13564: gap of 100 bp

13565 15205: contig of 1641 bp in length

15206 15305: gap of 100 bp

15306 19909: contig of 4604 bp in length

19910 20009: gap of 100 bp

20010 23963: contig of 3954 bp in length

23964 24063: gap of 100 bp

24064 29018: contig of 4955 bp in length

29019 29118: gap of 100 bp

29119 34619: contig of 5501 bp in length

34620 34719: gap of 100 bp

34720 42060: contig of 7341 bp in length

42061 42160: gap of 100 bp

FEATURES

source

42161 48254: contig of 6094 bp in length
 48255 48354: gap of 100 bp
 48355 55799: contig of 7445 bp in length
 55800 55899: gap of 100 bp
 55900 64293: contig of 8394 bp in length
 64294 64393: gap of 100 bp
 64394 70209: contig of 5816 bp in length
 70210 70309: gap of 100 bp
 70310 80175: contig of 9866 bp in length
 80176 80275: gap of 100 bp
 80276 90721: contig of 10446 bp in length
 90722 90821: gap of 100 bp
 90822 103432: contig of 12611 bp in length
 103433 103532: gap of 100 bp
 103533 116657: contig of 13125 bp in length
 116658 116757: gap of 100 bp
 116758 128699: contig of 11942 bp in length
 128700 128799: gap of 100 bp
 128800 143088: contig of 14289 bp in length
 143089 143188: gap of 100 bp
 143189 157620: contig of 14432 bp in length
 157621 157720: gap of 100 bp
 157721 174200: contig of 16480 bp in length.
 Location/Qualifiers
 1. 174200
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6"
 /clone="RP11-281G13"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 3194
 /note="assembly-fragment"
 3295 5794
 /note="assembly-fragment"
 5895 9861
 /note="assembly-fragment"
 9962 13464
 /note="assembly-fragment"
 13565 15205
 /note="assembly-fragment"
 15306 19909
 /note="assembly-fragment"
 19910 20009
 /note="assembly-fragment"
 20010 23963
 /note="assembly-fragment"
 24064 29018
 /note="assembly-fragment"
 29019 29118
 /note="assembly-fragment"
 29119 34619
 /note="assembly-fragment"
 34720 42060
 /note="assembly-fragment"
 42161 48254
 /note="assembly-fragment"
 48355 55799
 /note="assembly-fragment"
 55900 64293
 /note="assembly-fragment"
 64394 70209
 /note="assembly-fragment"
 70310 80175
 /note="assembly-fragment"
 80276 90721
 /note="assembly-fragment"
 90822 103432
 /note="assembly-fragment"
 103533 116657
 /note="assembly-fragment"
 116758 128699
 /note="assembly-fragment"
 128700 143088
 /note="assembly-fragment"
 143089 143188
 /note="assembly-fragment"
 143189 157620
 /note="assembly-fragment"
 157621 157720
 /note="assembly-fragment"

misc_feature 128800..143088 /note="assembly-fragment"
misc_feature 143189..157620 /note="assembly-fragment"
misc_feature 157721..174200 /note="assembly-fragment"
BASE COUNT 54534 a 31863 c 30617 g 55083 t 2103 others
ORIGIN

Query Match 1.88; Score 40; DB 2; Length 174200;
Best Local Similarity 58.3%; Pred. No. 3.6; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1897 tagtgccttaactcgcgtgtgttcacacatagatcatagaccttcattatcttcg 1956
DB 106406 TACTCTTCATTTAGCTTCTTCCACAAATGAATGACCTCTCTCTTTTAA 106465
QY 1957 ggccttgcttcagctacatgacacatagctagctcgtcacttgcgtgcgtatcgt 2016
DB 106466 ATTAAATGTTTTAGCTCAATATCTGTGACAGATCATTCCGACCTTGCGTGAATTGT 106525

RESULT 11
AF429315/c 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes/S.E., O'Hearn/E., Rosenblatt/A., Calzahan/C., Hwang/H.S.,
Ingersoll/Ashworth/R.G., Fleisher/A., Stevanin/G., Brice/A.,
Potter/N.T., Ross/C.A. and Margolis/R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes/S.E., Ingersoll-Ashworth/R.G., Ross/C.A. and Margolis/R.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES
source
1. 125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit=ctg
complement(<36507..>36887)
/gene="JPH3"
/product="Junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="JP3"
complement(<36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="Junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"

/translation="MSGGRFNEPDGSGYCGGMDGKAHGVCCTGPKCGEYTGSMSS
HGFEVLGVYTPSGNTYGGTWAQGRKRGIGLESKKGKVVYKKEWTGFKRGYVRECAAG
NGAKYEETGSMNLDGYCTETYSDG"
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN

Query Match 1.88; Score 38.6; DB 9; Length 125020;
Best Local Similarity 12.0%; Pred. No. 9.3; Indels 7; Gaps 4;
Matches 131; Conservative 449; Mismatches 506; Indels 7; Gaps 4;

QY 147 ctgtgatttggaggttcttgactatgagaagaggtccctgactgtgaatttgcga 206
DB 51857 SYMMCASTSRWKSAMSYWAGGTSTRKSGWGMKMTKGRSRKMSWMSKRSKMSW 51798
QY 207 atttagggaccacacgggttcttgagtccttgcacttatacaattgtgttac 266
DB 51797 RKGSRASARMSYSYCASRRKSCMRMRCGYMTGRSMGTMSKRSKRTSMGSKACTYR 51738
QY 267 aagctcgacgagttcactcgacgtagagtttagtcgggttcgaagatctatgcg 326
DB 51737 SARRAAMYSKSGRRRCKKSSKGGGRCGKKWMASTYSKRCACSTGRRGTKGSSC 51678
QY 327 ctgaagcggg--gaagagagcctgatttgcgaagcgatcgtgttgatttgcactaga 384
DB 51677 WTRRGMGTTRKAAMKCMCMWNYRGTGCMKATYATWYKMSKRWMTGRMRRA 51618
QY 385 gaggctgagccatcagaactatttctctctgtgaggaagatcgaagttccaatgg 444
DB 51617 WMMWMTWKKYRAHMKATRSMMASTYSKSTYARCCSMASAYKKYSRSHKCKMGK 51558
QY 445 tctactccgttcttctgtgtcgaagttcaagttatgctcgggtgttctaccccgag 504
DB 51557 MAASCRRTYRWKMGCGKCCSAGMTGRGS--RMSACMWSKRSASRARAAYATG 51500
QY 505 ccaagtcgagacatcgaacaatctgtagctgcgaagatctctctgtgttaactcaga 564
DB 51499 AAAMSGCYVKKMOTRRMNAATACARRRRAWMSCCMRKMTSKSKGRMRMYCMA 51440
QY 565 ctactatgtctcgtgtagcaacccttaccacaaatcgaatcgaatgtagcggaggaaggt 624
DB 51439 RMWYAGSYMSKCRSRMGYKGRSBSRSMSWRTRGYSARCYNKSAGAGCKAGS 51380
QY 625 gctcggatctgggtgtgactactcgttctgtgcgaagcgaatgcccacttcttaga 664
DB 51379 YAKSGWKMKGYSKMGAMYSKMSASKSRMTCTWSCCTSCGCTWMTWMSKYGAGMCM 51320
QY 685 acgtcagccgaagtagtacaatcgaagatcgaagatctgtccagatcttccaacatg 744
DB 51319 YMRMCTCMARAMRAARMAARARMAAGAAARMAWMSMMWMSMMRHAAMARMAAAK 51260
QY 745 ggtccatgtgcggaggtgtgcgaagccgacgaatggaagcgaatcgaagagttt 804
DB 51259 CMWNYSTRYCYMSWMAAAMAMMMCTAAWYKKWTRMRBMSYSSSAYAVATMC 51200
QY 805 tgaattcttcaagatcagatcagcagctt-gaatcatcagaacgaaggccagag 863
DB 51199 WYCYSSMYARMMYRSGKRRKMASTSCMCCRRGCMKKMKMRKCSYTTTGSGSKS 51140
QY 864 ctcaactgcgtggaagcgaatcgaagcgaatcgaatcgaatcgaatcgaatgc 923
DB 51139 WMTSSWACMYTKGSRKRTKSMATSGMNMGM--KSGKGMRTYCRCKMRMRNSGT 51082
QY 924 tgcgagacttgcactctagaatttcaatagaactatgaaatgcacatccatc 983
DB 51081 KCGSGTWSMGWGTGRTSKYSTGTGCKSCAMKMSKSTKSTSKYSCMGYVS 51022
QY 984 ttccggaagagcttctcgcgaagcgaatcgaagcgaatcgaatcgaatcgaatc 1043
DB 51021 YCTCMGWRMKMGKSSCCASRSCKSYKMGKSMRGRGRMDSMWSSMKSYSMAYSKS 50962
QY 1044 taaagatgcgcgaaggtcgaatggaatgtaatgttctggaaggaagaatctcact 1103

```

Db 50961 ARGGMKCYKTSCKMYYSMBHSSHDSBYSKSYBYWTCMKMKSMDVSHMRD 50902
Oy 1104 ttcgaagagcgctagagcgagcgttaagttcgtgaagccggtttcagagacatccg 1163
Oy 50901 MMYKMKMKYKSSSTGTYHBSHSMKMSDSDSGYSGMYHMKYKMSKMYSGMK 50842
Oy 1164 agatgaagcctgcatgcaatccacacatcgcagatgcgtgagatgggtctcgc 1223
Db 50841 SSVYBMCKMSMTYSBSKSYRMBRMSGTYBRYKCAVMKRYBHRSTRSVMTGKRSQV 50782
Oy 1224 cgaagaagcctc 1236
Db 50781 RDSVSRRRSVHTS 50769

RESULT 12
AC013612/c
ID AC013612 standard; DNM; HTG; 155542 BP.
XX
XX AC013612:
XX
XX AC013612.3
XX
XX 15-NOV-1999 (Rel. 61, Created)
XX 03-APR-2000 (Rel. 63, Last updated, Version 3)
XX Homo sapiens clone RP11-10L7, WORKING DRAFT SEQUENCE, 10 unordered pieces.
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
XX RP 1-155542
XX RA Birren B., Linton L., Nusbaum C., Lander E.:
XX "Homo sapiens, clone RP11-10L7";
XX Unpublished.
XX
XX
XX
XX [2]
XX RP 1-155542
XX RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
XX Baldwin J., Barna N., Beckerly R., Boguslavsky L., Boukhgalter B.,
XX Brown A., Castle A., Colangelo M., Collins S., Collymore A., Cooke P.,
XX Deatellano K., Dewar K., Domlino M., Donelan L., Doyle M., Ferreira P.,
XX Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyna S.,
XX Grant G., Hagos B., Heathford A., Horton L., Howland J.C., Johnson R.,
XX Jones C., Kann L., Karatas A., Klein J., Lehoczeky J., Lieu C., Locke K.,
XX MacDonald P., Marguis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
XX Meltrin J., Morrow J., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,
XX Peterson K., Pollara V., Riley R., Roy A., Santos R., Severy P.,
XX Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,
XX Tirrell A., Vassiliou H., Vo A., Wheeler J., Wu X., Wymann D., Ye W.Y.,
XX Zimmer A., Zody M.;
XX
XX Submitted (13-NOV-1999) to the EMBL/GenBank/DBJ databases.
XX Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
XX Cambridge, MA 02141, USA
XX
XX
XX On Mar 30, 2000 this sequence version replaced gi:6532098.
XX All repeats were identified using RepeatMasker:
XX Smit, A.F.A. & Green, P. (1996-1997)
XX http://ftp.genome.washington.edu/RM/RepeatMasker.html
XX
XX
XX Center: Whitehead Institute/ MIT Center for Genome Research
XX
XX Center code: WIBR
XX Web site: http://www-seq.wi.mit.edu
XX
XX Contact: sequence_submissions@genome.wi.mit.edu
XX
XX
XX Project Information
XX
XX Center project name: L3101
XX
XX Center clone name: 10_L_7
XX
XX
XX Summary Statistics

```

```

CC Sequencing vector: M13; M77815; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 151809 bases at least Q40
CC Consensus quality: 153486 bases at least Q30
CC Consensus quality: 154041 bases at least Q20
CC Insert size: 157000; agarose-fp
CC Insert size: 154642; sum-of-ctnigs
CC Quality coverage: 5.5 in Q20 bases; agarose-fp
CC Quality coverage: 5.6 in Q20 bases; sum-of-ctnigs
CC
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 10 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC
CC * 1 3918: contig of 3918 bp in length
CC * 3919 4018: gap of 100 bp
CC * 4019 7852: contig of 3834 bp in length
CC * 7853 7952: gap of 100 bp
CC * 7953 10420: contig of 2468 bp in length
CC * 10421 10520: gap of 100 bp
CC * 10521 16131: contig of 5611 bp in length
CC * 16132 16231: gap of 100 bp
CC * 16232 21713: contig of 5482 bp in length
CC * 21714 21813: gap of 100 bp
CC * 21814 25983: contig of 4170 bp in length
CC * 25984 26083: gap of 100 bp
CC * 26084 33604: contig of 7521 bp in length
CC * 33605 33704: gap of 100 bp
CC * 33705 45042: contig of 11338 bp in length
CC * 45043 45142: gap of 100 bp
CC * 45143 77339: contig of 32159 bp in length
CC * 77340 77439: gap of 100 bp
CC * 77440 155542: contig of 78103 bp in length.
CC
XX
XX Key Location/Qualifiers
XX FH
XX FT source 1. 155542
XX FT /db_xref="taxon:9606"
XX FT /organism="Homo sapiens"
XX FT /clone="RP11-10L7"
XX FT /clone_11b="RPC1-11 Human Male BAC"
XX FT 1. 3918
XX FT /note="assembly_fragment"
XX FT misc_feature 4019..7852
XX FT /note="assembly_fragment"
XX FT misc_feature 7953..10420
XX FT /note="assembly_fragment"
XX FT misc_feature 10521..16131
XX FT /note="assembly_fragment"
XX FT misc_feature 16232..21713
XX FT /note="assembly_fragment"
XX FT misc_feature 21814..25983
XX FT /note="assembly_fragment-clone_end:T7-vector_side:left"
XX FT misc_feature 26084..33604
XX FT /note="assembly_fragment"
XX FT misc_feature 33705..45042
XX FT /note="assembly_fragment"
XX FT misc_feature 45143..77339
XX FT /note="assembly_fragment-clone_end:SP6-vector_side:right"
XX FT misc_feature 77440..155542
XX FT /note="assembly_fragment"
XX
XX
XX Sequence 155542 BP; 44570 A; 32359 C; 30817 G; 46895 T; 901 other;
XX
XX
XX Query Match 1.8%; Score 38.6; DB 30; Length 155542;
XX Best Local Similarity 51.4%; Pred. No. 9.7;
XX Matches 89; Conservative 0; Mismatches 0; Gaps 0;

```

```

FEATURES
source
1
130323: Count of 1
Location/Qualifiers
1.156325
/organism="Homo sapiens"

```

```

Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated total Number of Errors is 0.9.
SIS Content:
SHGC-56762 G37043
SHGC-57254 G38378
WI-6292 G06202
SHGC-89248 G61707
WI-11361 G24462.
Location/Qualifiers
1..162914
format="Homo_sapiens"
SOURCE
```

FEATURES	
SOURCE	location/Qualifiers
	1..162914
	/organism:"Homo Sapiens"
	/db_xref:"taxon:9606"
	/chromosome:"19"
	/clone:"CTD-2102p23"
BASE COUNT	42170 a 36675 c 37407 g 46662 t
ORIGIN	

Oy 1168 gggctgctctatggaattctcaacatccgatatgcttaccggatagtggtctcccgaa 1227
 |||||
 Db 310 gaagcgtctatcgaaattctcaacatccgatatgcttaccggatagtggtctcccgaa 369
 |||||
 Oy 1228 gaagcgtctatcccaatctcttaccctggtcccgagaacattctaccgcatcaaatgag 1287
 |||||
 Db 370 gaagcgtctatcccaatctcttaccctggtcccgagaacattctaccgcatcaaatgag 429
 |||||
 Oy 1288 tgtcccgaaaggagagctctctccgacggagaagctctacacactctccacggccg -ca 1345
 |||||
 Db 430 tgtcccgaaaggagagctctctccgacggagaagctctacacactctccacggccgcat 489
 |||||
 Oy 1346 tctgttagcgcgagctgcg -tcgaagtattctctgtcatctccacgaagcttcaagaaatcg 1404
 |||||
 Db 490 tctgttagcgcgagctgcgagtatctcgtcgcacccagcaagtccaagaatcg 549
 |||||
 Oy 1405 a -ggcaagcgaaatctc -tctatagcgtagttgtgct -cgacgaactccatctat 1461
 |||||
 Db 550 agggcaagcgaaatctccttcttatatgtagtgattgtgcttcgcacgcggagctccatct 609
 |||||
 Oy 1462 attccccatccgataacatcgatctgcttgataaattctcccaactt 1510
 |||||
 Db 610 at -ccccatccgataaacty -atcgctggtatattctcccaattt 655
 |||||

FT			specification, and is included to maintain the
FT			base numbering given in the specification"
FT	misc-difference	3642	
FT		/*tag=	"h
FT		/note=	"this nucleotide is represented as a * in the
FT			specification, and is included to maintain the
FT			base numbering given in the specification"
FT	misc-difference	3657	
FT		/*tag=	i
FT		/note=	"this nucleotide is represented as a * in the
FT			specification, and is included to maintain the
FT			base numbering given in the specification"
FT	misc-difference	3659	
FT		/*tag=	j
FT		/note=	"this nucleotide is represented as a * in the
FT			specification, and is included to maintain the
FT			base numbering given in the specification"
FT	misc-difference	3681	
FT		/*tag=	k
FT		/note=	"this nucleotide is represented as a * in the
FT			specification, and is included to maintain the
FT			base numbering given in the specification"
FT	misc-difference	3698	

RESULT	/	3
AA84332/c		
ID	AA84332 standard; DNA; 5059 BP.	
XX	AA84332;	
XX		
D7	08-SEP-1999 (First entry)	
XX		
De	Stealth virus nucleic acid clone, SEQ ID NO: 24.	
XX		
KM	Stealth virus; detection; diagnosis; infection; ss.	
XX		
OS	Stealth virus.	
XX		
FH	Key	Location/Qualifiers
FT	misc-difference 3605	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3610	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3615	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3630	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3631	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3636	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	misc-difference 3641	/*tag=
FT		/*note=
FT		"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"

FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3703
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3708
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3713
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3720
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3732
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3733
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3740
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3747
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3750
FT	/*tag-	
FT	/note-	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	FT	
FT	FT	
FT	misc_difference	3751

[illegible][illegible]

PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229313
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230447
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234271
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0235484
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236357
PR	29-SEP-2000	2000US-0236358
PR	29-SEP-2000	2000US-0236359
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0242221
PR	01-NOV-2000	2000US-0246117
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526

XX	08-NOV-2000;	2000US-0246527.
XX	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246533.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249300.
PR	17-NOV-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	01-DEC-2000;	2000US-0251161.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251088.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-541565/60.	
XX		
XX	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system	
PT	cancers and metastases -	
XX		
PS	Disclosure; SEQ ID NO 7348; 1701ip + sequence listing; English.	
XX		
CC	The invention relates to novel genes (ABA11004-ABA21534) and proteins	
CC	(ABA1678-ABA18001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pcr_sequences.	
XX		
XX	Sequence 2625 BP; 699 A; 644 C; 581 G; 701 T; 0 other;	


```
RESULT 10
ABL32087
ID ABL32087 standard; DNA; 8178 BP.
XX
AC ABL32087;
XX
XX 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SFO ID NO: 60.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antileukocytic; antineoplastic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antileukemic; antileukemic; antidiabetic; antiparasitic;
KW antileukemic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX MO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001MO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIDEMIOLOGICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation.
XX
XX Claim 1: SEQ ID NO 60; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 8178 BP; 2164 A; 119 C; 1938 G; 3955 T; 2 other:
S0
Query Match 1.6%; Score 34; DB 24; Length 8178;
Best Local Similarity 52.9%; Pred. No. 11;
Matches 73; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
Oy 1934 tatgacctcctatcatttcgagcttgcttcgagctcattaccacgactagcgc 1993
Db 817 taagttatttctacgctgcttcaatttaattgatttcaatttcgtatagcgtt 876
Oy 1994 acttgaactgcgataatcgtcttgcaatttagtgagcttgatagatagcgc 2053
Db 877 atttttgaagagcttcaagaagatgatatttagtattgcttcaatgatagctgga 936
Oy 2054 atttgacagcagctgctgg 2071
Db 937 ttcttttggttgctgctgg 954
RESULT 11
```

```
AAF33260
ID AAF33260 standard; CDNA; 2354 BP.
XX
AC AAF33260;
XX
XX 23-MAR-2001 (first entry)
XX
DE Human secreted protein gene 48 SEQ ID NO:58.
XX
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antileukocytic;
KW dermatological; immunosuppressive; antineoplastic; anti-HIV;
KW immunostimulant; cytosolic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; noctropic; anticonvulsant; vulnary;
KW antileukemic; antiparkinsonian; antileukemic; immune disorder;
KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;
KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
KW cardiovascular disease; Schmitzer syndrome; Chaga's cardiomyopathy;
KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
KW corneal graft neovascularisation; neurological disorder; regeneration;
KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;
KW infectious disease; chemotaxis; ss.
XX
XX Homo sapiens.
XX
XX MO200076530-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000MO-US14933.
XX
XX 11-JUN-1999; 99US-0138572.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (ROSE-) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-071147/08.
XX
XX P-PSDB: AAB64929.
XX
XX Nucleic acids encoding 49 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy.
XX
XX Claim 1: Page 482-483; 554pp; English.
XX
XX The polynucleotide sequences given in AAF33213 to AAF33261 encode the
XX human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
XX AAB64991 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Examples of activities include:
XX immunomodulatory; antileukocytic; dermatological; immunosuppressive;
XX antineoplastic; anti-HIV; immunostimulant; cytosolic; cardiant;
XX vascular; antileukemic; anti-angiogenic; ophthalmological;
XX antiparkinsonian; anticonvulsant; noctropic; antileukemic;
XX antiparkinsonian; and vulnary. The polynucleotides and polypeptides can
XX be used in the prevention, diagnosis and treatment of diseases associated
XX with inappropriate polypeptide expression. Disorders that may be
XX prevented, diagnosed and/or treated by the above methods include immune
XX disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
XX human immunodeficiency virus (HIV) infections), hyperproliferative
XX disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
XX (e.g. Schmitzer syndrome, Chaga's cardiomyopathy and coronary
XX arteriosclerosis), angiogenic disorders (e.g. corneal graft
XX neovascularisation and diabetic retinopathy), neurological disorders
XX (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
XX infectious diseases and/or for promoting wound healing, regeneration and
XX /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
XX used in the exemplification of the present invention.
S0
Sequence 2354 BP; 606 A; 495 C; 525 G; 726 T; 2 other:
```


CC diseases. The present sequence is a gene of the invention.
XX Sequence 8486 BP: 1852 A; 263 C; 2416 G; 3955 T; 0 other;

Query Match 1.5%; Score 33.6; DB 24; Length 8486;
Best Local Similarity 55.0%; Pred. No. 15;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1647 ttgtgtccttgatcccgctgtgagagtgatcgtgcgcacccgtccatcagtggga 1706
DB 5061 ttgcgattccgagatcgcttctttagagacgagagcgggcgctgttcggt 5120
QY 1707 ttctctgagcagatctgtctcgcagttgttcataacgtttgcgtagaagg 1766
DB 5121 ttgattgcgagagattcgttctgatttctgtgcgcgttcctgcgtcgtg 5180

RESULT 14
AAAL0595/c
ID AAAL0595 standard; DNA: 6741 BP.

AC AAAL0595;

DT 29-JUN-2000 (first entry)

DE Gene encoding a subunit of cellulose synthase.

KW Cellulose synthase; cellulose production; increase yield; ds.

OS Vigna angularis.

PN Jp2000060568-A.

PD 29-FEB-2000.

XX 26-AUG-1998; 98JP-0239998.

PR 26-AUG-1998; 98JP-0239998.

XX (MIZU) MIZUNO K.

PA (OJIP) OJI PAPER CO.

XX WPI: 2000-342371/30.

DR P-PSDB: AAY85180.

PT A gene encoding a cellulose synthetic equipment - for the improvement

PT in the amount of cellulose synthesised in a plant body

PS Claim 2; Page 22-31; 32pp; Japanese.

XX This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.

XX Sequence 6741 BP: 1712 A; 870 C; 1468 G; 1374 T; 1317 other;

Query Match 1.5%; Score 33.4; DB 21; Length 6741;
Best Local Similarity 17.6%; Pred. No. 15; Mismatches 210; Indels 2; Gaps 1;
Matches 78; Conservative 154;

QY 78 ctgagagtcgcagcagaagtaactgacctgcgcgaatctggaattggaagcgctg 137

DB 6135 CSRKAKSYNCCDDYTCNSKCNSTCCYCYSSRSRNCTTSRRCTBYRAVACSRSSTSDSKSR 6076

QY 138 aagctcacctctgattgagagtcgtctgcacgatgagagaggtctcagctgagga 197

DB 6075 RCMSTTTCBRTCYTTTSTTTSTNSTCDNSTBDRCDSDGSRYSRRCSTRCSTCTA 6016

QY 198 atgttgcaatttaggggacacacggttctgtgagtccttcgcaacttaaccaat- 256
DB 6015 KTBETTYSYDAYARCDRCRDYASRCSSTSYDAYSSTTNSRTBDBNSTMD 5956

QY 257 -gttggttacaagctcgcagagttcaatcagcagatgagttttagctgagtgagat 315

DB 5955 SRTTSRSTTTTADCCYTCNCSRCYTDYCTBYRATBNCSSRSTSTAKRNSNCTTSRN 5896

QY 316 ctatgtatccgcctcagcgagagagagcccgatgtgcgcgaagcagatcggtgatt 375

DB 5895 CTTTSCCNSTPDCBTRCYNRACSRYSNCRYSYSTYDATTBBSRGYSNSTKDY 5836

QY 376 gactagaagagtgagcagcagcatcagaactattatccttgcgagggagagatcagat 435

DB 5835 DAVSDYDASTDTSRSDSTTTCYTSTNSRDNSTRCTTRCYSRCCYYSYCTSDY 5776

QY 436 tccaatgggtctcaactcgttctgtgtgtagcaggtcaggtatgctcgggtgctta 495

DB 5775 SYRAKCYTYSRNSTNSTYTDADNSTBTSITTCYTTTSTNCNNTYSYRANSTTBXRA 5716

QY 496 ccccgagcgccagctcagagaatct 519

DB 5715 RCYRACYSCTYCTYSYDASTCY 5692

RESULT 15

ABL3307B
ID ABL3307B standard; DNA: 8375 BP.

AC ABL3307B;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1051.

KW Human; immune system disease; cytosine methylation; antileukemic;

KW antiarteriosclerotic; antianemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antiparasitic;

KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001MO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPICENOMICS AG.

PI Olek A. Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -
PS Claim 1; SEQ ID NO 1051; 32pp + sequence listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:47:06 : Search time 59.1 Seconds
(without alignments)
9048.124 Million cell updates/sec

Title: US-09-828-313-22

Sequence: 2177
1 atccggagctgtatgtgct.....gcagatatgccgttaacgc 2177

Scoring table: IDENTITY NUC
Gapop 10.0, Gapept 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgml2/6/ptodata/2/ina/5A.COMB.seq:*
2: /cgml2/6/ptodata/2/ina/5B.COMB.seq:*
3: /cgml2/6/ptodata/2/ina/6A.COMB.seq:*
4: /cgml2/6/ptodata/2/ina/6B.COMB.seq:*
5: /cgml2/6/ptodata/2/ina/PCUS.COMB.seq:*
6: /cgml2/6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41.6	1.9	7218	1	US-08-232-463-14 Sequence 14, Appl
2	31.8	1.5	1935	2	US-08-878-989-11 Sequence 11, Appl
3	31.8	1.5	1935	4	US-09-272-796-11 Sequence 11, Appl
4	31.8	1.5	4403765	4	US-09-103-840A-2 Sequence 2, Appl
5	31.8	1.5	4411529	4	US-09-103-840A-1 Sequence 1, Appl
6	31.2	1.4	396	3	US-08-963-168C-4 Sequence 4, Appl
7	31.2	1.4	458	3	US-09-141-000-4 Sequence 4, Appl
8	30.4	1.4	1101	1	US-09-078-294-20 Sequence 20, Appl
9	30.4	1.4	1791	1	US-08-245-284-7 Sequence 7, Appl
10	30.4	1.4	1791	1	US-08-474-499-7 Sequence 7, Appl
11	30.4	1.4	1791	5	US-08-307-279A-7 Sequence 7, Appl
12	30.4	1.4	1791	5	PCT-US95-06211-7 Sequence 7, Appl
13	30.4	1.4	4164	1	US-08-204-675-1 Sequence 1, Appl
14	30.4	1.4	4164	2	US-08-660-754-1 Sequence 1, Appl
15	30.4	1.4	4164	2	US-08-796-364-1 Sequence 1, Appl
16	30.4	1.4	4164	5	PCT-US95-02520-1 Sequence 1, Appl
17	30.4	1.4	4933	1	US-08-304-675-2 Sequence 2, Appl
18	30.4	1.4	4933	2	US-08-660-754-2 Sequence 2, Appl
19	30.4	1.4	4933	2	US-08-796-364-2 Sequence 2, Appl
20	30.4	1.4	4933	5	PCT-US95-02520-2 Sequence 2, Appl
21	30.4	1.4	5181	3	US-08-801-344-5 Sequence 5, Appl
22	30.4	1.4	5181	4	US-09-498-599-5 Sequence 5, Appl
23	30.4	1.4	7102	3	US-09-138-024-20 Sequence 20, Appl
24	30.4	1.4	7333	3	US-09-138-024-21 Sequence 21, Appl
25	30.4	1.4	7633	3	US-09-028-851-1 Sequence 1, Appl
26	30.4	1.4	7633	3	US-08-815-520-1 Sequence 1, Appl
27	30.4	1.4	7633	4	US-09-273-163-1 Sequence 1, Appl

28	30.4	1.4	9936	4	US-08-972-927-2 Sequence 2, Appl
29	30	1.4	2933	4	US-08-936-165A-201 Sequence 201, App
30	30	1.4	5163	3	US-08-700-651-1 Sequence 1, Appl
31	30	1.4	5163	3	US-08-928-361B-4 Sequence 4, Appl
32	30	1.4	5318	3	US-08-700-651-2 Sequence 2, Appl
33	30	1.4	5318	3	US-08-928-361B-3 Sequence 3, Appl
34	30	1.4	5318	3	US-08-928-361B-2 Sequence 2, Appl
35	30	1.4	7334	3	US-08-928-361B-1 Sequence 1, Appl
36	29.8	1.4	1342	3	US-08-445-515-54 Sequence 54, Appl
37	29.8	1.4	1558	1	US-08-467-607-2 Sequence 2, Appl
38	29.8	1.4	1558	2	US-08-469-362-2 Sequence 2, Appl
39	29.8	1.4	1558	2	US-08-850-392-2 Sequence 2, Appl
40	29.8	1.4	6253	3	US-08-893-327-15 Sequence 15, Appl
41	29.8	1.4	6280	3	US-08-893-327-17 Sequence 17, Appl
42	29.8	1.4	6280	3	US-08-893-327-19 Sequence 19, Appl
43	29.6	1.4	2166	2	US-08-317-401E-3 Sequence 3, Appl
44	29.6	1.4	2877	2	US-08-317-401E-1 Sequence 1, Appl
45	29.6	1.4	3591	1	US-07-943-843-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232.463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
STRANDEDNESS: single
IMMEDIATE SOURCE: F15
CLONE: pTZap1-Fls
US-08-232-463-14


```
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SYNBRAT04
CLONE: 705365
US-09-272-796-11
```

```
Query Match 1.5%; Score 31.8; DB 4; Length 1935;
Best Local Similarity 51.8%; Pred. No. 2.7;
Matches 72; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
```

```
Y 1489 ttggaataattcccaacattttgctcttcaccccaagagcctcctaagcgacgt 1548
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1593 ttgagaaatccatctcctgtagaatttcccttggtagcattgacaccacattcttagg 1534
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 1549 aatagttaacgacattgacaaactcccaattacgtagcgttattctgtaaccagcttc 1608
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1533 actaatttcagtattgttatctacgccaagcagcctgtgtatgcactgactgagtttc 1474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 1609 atcgaagtatacgaagatg 1627
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 1473 tcttggttatggggaatg 1455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4

```
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FRATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
```

```
Query Match 1.5%; Score 31.8; DB 4; Length 4403765;
Best Local Similarity 52.7%; Pred. No. 54;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
Y 817 agtaacgatgcagcagcttgaatcatacgaagagggccagagcctcctcgcttg 876
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4337020 AGTGGCTTGCAACGAGCTAGTCCGCTCGACCAATTTGGGAGGCGAAGAGCTCGGA 4326961
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 877 agaccgacatcgaagacacttaagctcgtttcatcgcgcattcgttcgagacactcg 936
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4326960 TGCCCGACCATCGAATCGTCAATGGGAGCATTTGGCAATTAAGAGAGACTCA 4326901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 937 acttctcagaa 947
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4326900 ACCTGCTGGTA 4326890
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 5
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
```

```
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 1.5%; Score 31.8; DB 4; Length 4411529;
Best Local Similarity 52.7%; Pred. No. 54;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
Y 817 agtaacgatgcagcagcttgaatcatacgaagagggccagagcctcactcgcttg 876
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4334766 AGTGGCTTGCAACGAGCTAGTCCGCTCGACGAATTTGGGAGGCGAAGAGCTCGGA 4334707
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 877 agaccgacatcgaagacacttaagctcgtttcatcgcgcattcgttcgagacactcg 936
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4334706 TGCCCGACCATCGAATCGTCAATGGGAGCATTTGGCAATTAAGAGAGACTCA 4334647
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 937 acttctcagaa 947
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4334646 ACCTGCTGGTA 4334636
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6

```
US-08-963-168C-4/c
Sequence 4, Application US/08963168C
Patent No. 6127166
GENERAL INFORMATION:
APPLICANT: Bayley, Hagan
APPLICANT: Cao, Yunping
APPLICANT: Wang, Yunjuan
TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES
TITLE OF INVENTION: AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
```

```
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: PASTESQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,168C
FILING DATE: 03-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/059001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
```


CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.612
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 141..1649
US-08-245-294-7

Query Match 1.4% Score 30.4; DB 1; Length 1791;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccatcgagatactgcatcgttgataaattctccacaattttgctct 1518
DB 1390 TCCAAATGTTTCATCTGAATTCACCGCTCATTTGAATATTGAGCCATCTTCATATT 1331
QY 1519 tcaaccacagcagctccctcaatgagcaataatgtagacatttgacaactccaa 1578
DB 1330 TCGCTTCAGGATCGAATCAAGTTAACTTAATTTCTCTTACCGATTTCCTCA 1271
QY 1579 ttaagtagcgttatctgtaaccacgttcat 1610
DB 1270 ACTCCTAAGTTACTGTTCTCTGCGCTCAT 1239

RESULT 10
US-08-474-499-7/C
Sequence 7, Application US/08474499
Patent No. 5693776
GENERAL INFORMATION:
APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,499
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,294
FILING DATE: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.612
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 141..1649
US-08-474-499-7

Query Match 1.4% Score 30.4; DB 1; Length 1791;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctattcccatcgagatactgcatcgttgataaattctccacaattttgctct 1518
DB 1390 TCCAAATGTTTCATCTGAATTCACCGCTCATTTGAATATTGAGCCATCTTCATATT 1331
QY 1519 tcaaccacagcagctccctcaatgagcaataatgtagacatttgacaactccaa 1578
DB 1330 TCGCTTCAGGATCGAATCAAGTTAACTTAATTTCTCTTACCGATTTCCTCA 1271
QY 1579 ttaagtagcgttatctgtaaccacgttcat 1610
DB 1270 ACTCCTAAGTTACTGTTCTCTGCGCTCAT 1239

RESULT 11
US-08-307-279A-7/C
Sequence 7, Application US/08307279A
Patent No. 5736347
GENERAL INFORMATION:
APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
TITLE OF INVENTION: Henselae and Rochalimaea Quintana Infection
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, N.E., Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,279A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 688-0770
TELEFAX: (404) 688-9880
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 141..1652
US-08-307-279A-7

Query Match 1.4%; Score 30.4; DB 1; Length 1791;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctatcccccacagatgacgacgttcgttgataaattctccacaattttgctct 1518
DB 1330 TCCAAATGTTTCATCTGAATTCACGCTCATTTGAATTTTGAGCCATCTTCATATTT 1331
QY 1519 tcaccctcaagaagctcctcaatgcccgaataatgtagaactgtagcaactccaa 1578
DB 1330 TCGCTTCAGGCGATCGAATTCAGTTTAACCTTAATTCCTCTTACCAAGATTTCCA 1271
QY 1579 ttaagtagctatctcgttaaccacagttcat 1610
DB 1270 ACTCCTAAGGTTACTGTTTCTCTCGGGCTCAT 1239

RESULT 12
PCT-US95-06211-7/c
Sequence 7, Application PC/TUS9506211

GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06211

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,294

FILING DATE: 18 MAY 1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.6121

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 141..1649

PCT-US95-06211-7

Query Match 1.4%; Score 30.4; DB 5; Length 1791;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1459 tctatcccccacagatgacgacgttcgttgataaattctccacaattttgctct 1518
DB 1330 TCCAAATGTTTCATCTGAATTCACGCTCATTTGAATTTTGAGCCATCTTCATATTT 1331
QY 1519 tcaccctcaagaagctcctcaatgcccgaataatgtagaactgtagcaactccaa 1578
DB 1330 TCGCTTCAGGCGATCGAATTCAGTTTAACCTTAATTCCTCTTACCAAGATTTCCA 1271
QY 1579 ttaagtagctatctcgttaaccacagttcat 1610
DB 1270 ACTCCTAAGGTTACTGTTTCTCTCGGGCTCAT 1239

RESULT 13
US-08-204-675-1
Sequence 1, Application US/08204675

Patient No. 5677170

GENERAL INFORMATION:
APPLICANT: Devine, Scott E.

APPLICANT: Boeke, Jef D.

APPLICANT: Bralderman, Lelita T.

TITLE OF INVENTION: In Vitro Transposition of Artificial

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie, and Beckett

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/204,675

FILING DATE: 02-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 37,141

REFERENCE/DOCKET NUMBER: 01107.45501

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202.508.9100

TELEFAX: 202.508.9289

TELEX: 197430 BMB UT

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4164 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: PAT-1
US-08-204-675-1

Query Match 1.4%; Score 30.4; DB 1; Length 4164;
Best Local Similarity 55.8%; Pred. No. 14;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2064 taagctgtagctgtttacaagaatagcgacagctagaagccttaaaccttagctacat 2123
DB 605 TATCTGACTGATTTTCCATGAGGCGACAGTTTAAGCCGCTTAAGGCTTATCGGCCAA 664
QY 2124 gttatttaaacacatagatgaagcgctgtgcaaatatg 2167
DB 665 GTACAATTTTACTCTTCGAAGACAGAAATTTCTGACATTG 708

```

RESULT 14
US-08-660-754-1
; Sequence 1, Application US/08660754
; Patent No. 5843772
; GENERAL INFORMATION:
; APPLICANT: Devine, Scott E.
; APPLICANT: Boeke, Jef D.
; APPLICANT: Braiterman, Leilita T.
; TITLE OF INVENTION: In Vitro Transposition of Artificial
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie, and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,754
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/204,675
; FILING DATE: 02-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.45501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: PAT-1
;
US-08-660-754-1

Query Match 1.4%; Score 30.4; DB 2; Length 4164;
Best Local Similarity 55.8%; Pred. No. 14;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2064 tagcttgagtcgctgtttacaagatacgagcagctagaagccttaaccttagctacat 2123
DB 605 TATCTTGACAGATTTCCTCATGAGGCGACACTTAAGCGCTAAAGCATTTATCCGCCAA 664

QY 2124 gttatttaaacctataatgaatgaacgagctgtgcagatatg 2167
DB 665 GTACAATTTTCTCTTCGAGACAGAAAATTTCTGACATTTG 708

RESULT 15
US-08-796-364-1
; Sequence 1, Application US/08796364
; Patent No. 5986785
; GENERAL INFORMATION:
; APPLICANT: Devine, Scott E.

```

```

; APPLICANT: Boeke, Jef D.
; APPLICANT: Braiterman, Leilita T.
; TITLE OF INVENTION: In Vitro Transposition of Artificial
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie, and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,364
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/204,675
; FILING DATE: 02-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.45501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: PAT-1
;
US-08-796-364-1

Query Match 1.4%; Score 30.4; DB 2; Length 4164;
Best Local Similarity 55.8%; Pred. No. 14;
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2064 tagcttgagtcgctgtttacaagatacgagcagctagaagccttaaccttagctacat 2123
DB 605 TATCTTGACAGATTTCCTCATGAGGCGACACTTAAGCGCTAAAGCATTTATCCGCCAA 664

QY 2124 gttatttaaacctataatgaatgaacgagctgtgcagatatg 2167
DB 665 GTACAATTTTCTCTTCGAGACAGAAAATTTCTGACATTTG 708

```

Search completed: September 16, 2002, 01:17:18
Job time: 12612 sec

KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera. Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 855)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT	<p>Web : www.genoscope.cns.fr</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..855</p> <p>/organism="Drosophila melanogaster"</p> <p>/plasmid="pBelobAC11"</p> <p>/db_xref="taxon:7227"</p> <p>/clone_lib="DrosBAC"</p> <p>/clone="BACN08k20"</p> <p>/note="end : Sp6"</p>
BASE COUNT	96 a 63 c 75 g 160 t 461 others
ORIGIN	
Query Match	1.9% ; Score 42; DB 12; Length 855;
Best Local Similarity	12.1% ; Pied. No. 0.4;
Matches	65; Conservative 129; Mismatches 344; Indels 0; Gaps 0;
Db 1522	tctctaaagcagctccctaattgcccagtaattatgtaagcaattgtgcacaactccaatta 1581
Db 694	kbbkkaakcocccknnknnknnknnknnknnknnknnknnknnknnknnknnknn 635
Db 1582	cgtgaagctatctcgttaaccacgctatcagatcagaagaatgctgcagtaagaact 1641
Db 634	nnkkkknknnknnknnknnknnknnknnknnknnknnknnknnknnknnknnknn 575
Db 1642	gctactctgctctgtgctacccgctgtgtgacgaagatcaltgctgcacccgtgcattcag 1701
Db 574	nnnnnnnnknnknnknnknnknnknnknnknnknnknnknnknnknnknnknnknn 515
Db 1702	tggagattctcttgagcagacagcttgcttcgcagattgtgttccaataagcttggctgt 1761
Db 514	knnkkkkknnnnnnnnknnknnknnknnknnknnknnknnknnknnknnknnkng 455
Db 1762	agggagccctagacgctactctcaagaacatgagaagtgtgtgtgtgtgtgtgtgtgt 1821
Db 454	nkgcgkknkkgkknknnknnknnnnnnnnnnnnnnnnnnnnknnknnknnknnknn 395
Db 1822	tctcttgagagatctgtcttccctatctagaacataagcagagacactgtgcctgtggcg 1881
Db 394	nnknnkktgganagknnnannanakaakkaakngkcccmkcmkcmkmaacnnnncnkn 335
Db 1882	acatccatagaaacatagctctcaactcttggtgtgtgttcaaccatagatcatgacct 1941
Db 334	nnnknknnnnnnknnknnknnknnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 275
Db 1942	tctcaatctcttgagcactgtcttgctagacatgacatgacacatgacatgacatgac 2001
Db 274	kkkkkkktgttngcckkkkkkkkkkkkkkkgagacagaaannnnkntnnnannnnangn 215
Db 2002	tgcgtgataacgtctgtccaatttagtgagacttgtgtagatgacatgacatgacatgtgc 2059
Db 214	tgaagagcttcttttgaamaakcmatttctgttttaragagaagcttctgtattattttt 157

RESULT	6	AG029463	949 bp	DNA	Linear	GSS 01-NOV-2001
LOCUS		AG029463/c				
DEFINITION		pan troglodytes DNA, clone: PTB-001122.F, genomic survey sequence.				
ACCESSION		AG029463				
VERSION		AG029463.1	GI:1655636			
KEYWORDS		GSS: GSS (genome survey sequence).				
SOURCE		pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male				
ORGANISM		BAC Library clone:PTB-001122.F.				
REFERENCE		pan troglodytes				
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.				
TITLE		1 (sites)				
JOURNAL		Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
REFERENCE		Totoki, Y., Watanabe, H., and Sakaki, Y.				
AUTHORS		BAC end sequences of library PTB				
TITLE		Unpublished				
JOURNAL		2 (bases 1 to 949)				
REFERENCE		Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
AUTHORS		Totoki, Y., Watanabe, H., and Sakaki, Y.				
TITLE		Submitted Submission				
JOURNAL		Direct Submission				
REFERENCE		Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical				
AUTHORS		and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
TITLE		1-7-22 Suehiro-chou Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
JOURNAL		(E-mail:chibmpbs@gsc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,				
REFERENCE		Tel:81-45-503-9111, Fax:81-45-503-9170)				
AUTHORS		Clones are derived from the chimpanzee BAC library PTB this BAC end				
TITLE		was generated during the R&D process and may have higher chance of				
JOURNAL		clone tracking errors.				
REFERENCE		PRIMERS				
AUTHORS		Sequencing: -21M13				
TITLE		LIBRARY				
JOURNAL		Vector 1 : pKS145				
REFERENCE		R.Site 1 : SacI				
AUTHORS		R.Site 2 : Sact.				
TITLE		Location/Qualifiers				
JOURNAL		1 : 949				
REFERENCE		/organism="Pan troglodytes"				
AUTHORS		/db_xref="taxon:9598"				
TITLE		/clone="PTB-001122.F"				
JOURNAL		/sex="male"				
REFERENCE		/cell_type="lymphoblast"				
AUTHORS		/clone_lib="PTB Chimpanzee Male BAC library"				
TITLE		325 a 311 c 100 g 208 t 5 others				
JOURNAL		BASE COUNT				
REFERENCE		ORIGIN				
AUTHORS		Query Match				
TITLE		Best Local Similarity 59.18; Pred. No. 2;				
JOURNAL		Matches 68; Conservative 0; Mismatches 47; Indels 0; Gaps 0;				
REFERENCE		QY 148 tgcgatttcggaggttcctgcactgcagtgagagagtcctcgtgcgtgagaaatgttcgaa 207				
AUTHORS		Db 494 tgcgatttcggaggttcctgcactgcagtgagagagtcctcgtgcgtgagaaatgttcg 435				
TITLE		OY 208 tttagagagacacacggcttcgttcgagagcccttcgacactacaacatgctgtg 262				
JOURNAL		Db 434 tttagagagagacacggcttcgttcgagagcccttcgacactacaacatgctgtg 380				
REFERENCE		RESULT 7				
AUTHORS		CNS00100/c				
TITLE		LOCUS				
JOURNAL		DEFINITION				
REFERENCE		CNS00100				
AUTHORS		BAC323D23 of RCI-98 library from Drosohilla melanogaster (fruit				
TITLE		fly), genomic survey sequence.				
JOURNAL		AL068607				
REFERENCE		AL068607.1	GI:4958669			
AUTHORS		GSS:				
TITLE		Fruit fly.				
JOURNAL		Drosophila melanogaster				
REFERENCE		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
AUTHORS		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				

REFERENCE 1 (bases 1 to 1101) Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT The determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mosmoser in Piere de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCT-98"
/clone="BACR32D23"
/note="end : TERT3"

BASE COUNT 155 a 166 c 7 g 284 t 489 others
ORIGIN

Query Match 1.8%; Score 39.2; DB 12; Length 1101;
Best Local Similarity 18.4%; Pred. No. 3.5; Mismatches 155; Indels 0; Gaps 0;
Matches 64; Conservative 129; Mismatches 155; Indels 0; Gaps 0;

QY 148 tggatattggaggtgtgttgcacgtagaagaaggtctgcagcctagaagtatttgcga 207
DB 1095 TTTTATKTKKKGKTKGTATKTDKNAKVAAKARTDDKTKATGAKGKKKDAKAA 1036
QY 208 tttagggagcaccggttltgtgagtcctgcacacttaacaattgtgttaca 267
DB 1035 KTKTKGMAAAMAAKTKHKTDTKGAKTKTKGKTGKKCKTKGKGRKDAKAA 976
QY 268 agctcagcagattcaatcgaagctttagtcgggtcggagatcattatccgc 327
DB 975 AADDKTKDRAKAKKGGKKKKKKKTKKTKGKMDDDAMMDMADTD 916
QY 328 tcagcgagagaagagcctgatgttcgcaagcgatcgltgtgatttagaagaag 387
DB 915 WKKAGGGGKGDADTDTWMDKAMKGGDKDKTKKKKGGGKAGRAAARAKAAD 856
QY 388 gtggaccgacatcaactattatcctgttgaggaagagatcgaagttcaatgttct 447
DB 855 GDAAKTKTKAKTKMKGKKGGGMRKGTGKRGKGTAKADTKKKKGRKKKGTGD 796
QY 448 caccctgttttctgtgcacggtcgaaggtatgtccgggtgtctca 495
DB 795 KKRtWMTTKTKTKGTAKTKTGKKKKKGGKKGKGDWTKGTKMD 748

RESULT 8
AM625386 553 bp mRNA linear EST 18-MAY-2001
LOCUS AM625386
DEFINITION EST131209 tomato radicle, 5 d post-imbibition, Cornell University
ACCESSION AM625386
VERSION AM625386.1 GI:7338329
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE 1 (bases 1 to 553) Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

AUTHORS van der Hoeven, R.S., Garvin, D., Matern, A.L., Holt, I.E., Liang, F., Hansen, T., Craven, M.B., Bowman, C.L., Ronling, C.M., Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
TITLE Generation of ESTs from tomato radicle tissue (etolated)
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source 1..553
/organism="Lycopersicon esculentum"
/cultivar="7A496"
/db_xref="taxon:4081"
/clone="cLEF13E15"
/clone_lib="tomato radicle, 5 d post-imbibition, Cornell University"
/tissue_type="radicle"
/dev_stage="seedlings 5 days post-imbibition"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Tankley; Tissue supplied by Dave Garvin (USDA-ARS, Ithaca, NY 14850)."

BASE COUNT 132 a 157 c 125 g 139 t
ORIGIN

Query Match 1.8%; Score 38.6; DB 9; Length 553;
Best Local Similarity 52.9%; Pred. No. 3.2; Mismatches 74; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1233 ccttaccatcttaccctcgtccgcgaagaattcacccgcacgacagatgttcc 1292
DB 307 CATCCCCAGCACACGAGACGACGCTCCGAGACTTCTACACACGACGACGCCCTGCTT 366
QY 1293 cgcgaagagacagtcctccgcgcgaagaagctcacacccctccagccgcatgttga 1352
DB 367 CGGAAGCAGCGTCCTCAATCATGATGTCTCCCTCCTCCACGACTTCACTACAG 426
QY 1353 cgcgcagctgcgtcgaagattctgcctgcattcagca 1389
DB 427 CTGGTTCTCCGCGGATGATGTCACGATTCGGA 463

RESULT 9
AM930530 671 bp mRNA linear EST 18-MAY-2001
LOCUS AM930530
DEFINITION EST140987 tomato fruit mature green, TAMU Lycopersicon esculentum
ACCESSION CDNA clone cLEF5063 5', mRNA sequence.
VERSION AM930530
KEYWORDS AM930530.1 GI:8105931
EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 671)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Alm, S., Ronling, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA


```
HVCDNA0010 (20 DAP)"
/tissue_type="20 DAP spike"
/lab_host="corn"
```



```

REFERENCE      2 (bases 1 to 1067)
AUTHORS        Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
                Bernot,A., Fizames,C., Wincker,P., Brotlier,P., Quetier,F.,
                Saurin,W. and Weissenbach,J.
TITLE          Human gene number estimate provided by genome wide analysis using
                Tetraodon nigroviridis DNA sequence
JOURNAL        Unpublished
REFERENCE      3 (bases 1 to 1067)
AUTHORS        Genoscope.
TITLE          Direct Submission
COMMENT        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
                This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the Tetraodon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/Tetraodon.

FEATURES
  source       1..1067
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="021D17"
                /clone_1lb="G"
                /note="Genoscope sequence ID : C08G021CB09LP1-end : T7"

BASE COUNT    356 a      185 c      268 g      257 t      1 others

ORIGIN

```

Query Match 1.7%; Score 37.2; DB 12; Length 1067;
 Best Local Similarity 54.3%; Pred. No. 14;
 Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

OY 1702 tgggatttccttgagcgacagatcgtcgcgagttgttcataagcttggtcgt 1761
      || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 TGATTTATGTTGACTGCACTGTGTGCGTCCGCCGCTTTTAAATTTGATTGCT 565

OY 1762 agggggccttagacgagctacatacagcaatgagaagtgtgctggtggaattgacagcaa 1821
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 GGGGGTCAGTCACTTCTTCTCAAGCAATGATACATGTTTGTGTGTCATATATAGAA 505

OY 1822 tccttgagagatgtct 1839
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 TAGTTTGTGTAAAGTTT 487

```

Search completed: September 15, 2002, 22:53:08
 Job time: 5077 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2002, 23:46:11 ; Search time 58.43 Seconds
(without alignments)
629.222 Million cell updates/sec

Title: US-09-828-313-35
Sequence: 1734
1 MGLTFPSCVVGQYRVVYP.....SAASQGVQNRNKAISLYMA 331

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	7.2	1381	22	ABR70001
2	116	6.7	912	16	ABR85442
3	105.5	6.1	2001	22	ABR20062
4	105.5	6.1	2001	22	ABR20063
5	105	6.0	560	22	AAU22961
6	104.5	6.0	377	22	ABR03643
7	104	6.0	1284	22	ABR58814
8	103.5	6.0	750	12	AA15150
9	103.5	6.0	1321	17	ABR36994
10	103	5.9	2742	21	ABR23012
11	103	5.9	2783	13	ABR23962

12	103	5.9	2783	13	ABR23963	AFP-1 (Ala 2460 Va
13	103	5.9	2783	22	ABR82946	Human androgen rec
14	103	5.9	2842	15	ABR63508	Adenomatous polypo
15	103	5.9	2843	13	ABR26052	APC gene product 1
16	103	5.9	2843	15	ABR58634	Adenomatous polypo
17	103	5.9	2843	16	AAW11972	Adenomatous polypo
18	103	5.9	2843	18	AAW35392	Human adenomatous
19	103	5.9	2843	19	AAW6140	Human APC protein
20	103	5.9	2843	19	AAW38370	Human APC protein
21	103	5.9	2843	15	ABR33011	Human APC protein
22	103	5.9	2860	21	ABR3507	Adenomatous polypo
23	103	5.9	2973	19	AAW70304	Human APC protein
24	103	5.9	2973	22	AAW72782	Transcriptional ac
25	103	5.9	2973	22	ABR59466	Drosophila melanog
26	103	5.9	1231	22	ABR58808	Drosophila melanog
27	102	5.9	795	22	ABR68471	Drosophila melanog
28	101.5	5.9	795	22	AAW44247	Human cell signal
29	101.5	5.8	519	19	AAW61739	M. tuberculosis im
30	101	5.8	679	19	AAW64372	M. tuberculosis an
31	101	5.8	679	20	AAW39026	M. tuberculosis re
32	101	5.8	679	20	ABR61913	Drosophila melanog
33	101	5.8	679	20	ABR57448	Mouse YNK1 partial
34	101	5.8	382	21	ABR86041	Drosophila melanog
35	100	5.8	611	22	ABR29283	Arabidopsis thalia
36	100	5.8	632	21	AAW29282	Arabidopsis thalia
37	100	5.8	734	21	AAW29281	Human WART1, ortho
38	100	5.8	736	21	AAW6526	Arabidopsis thalia
39	100	5.8	1130	20	AAW36453	Arabidopsis thalia
40	100	5.8	1222	21	AAW36452	Arabidopsis thalia
41	100	5.8	1275	21	AAW36451	Arabidopsis thalia
42	100	5.8	1275	21	AAW36451	Human protein sequ
43	100	5.8	1275	21	AAW36451	
44	100	5.8	1275	21	AAW36451	
45	99.5	5.7	285	22	ABR94278	

ALIGNMENTS

RESULT 1	
ABR70001	standard; Protein; 1381 AA.
ID	ABR70001
AC	ABR70001
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide seq ID NO 36795.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	NO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-0509231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PMD, Myers EW,
XX	
DR	WPI: 2001-656860/75.
XX	N-PSDB; ABL14104.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

FT	Region	177..350	/note="Internal repeat region"
FT	Peptide	362..367	/note="nuclear localization motif"
FT	Binding-site	460..467	/note="ATP/GTP-binding motif"
FT	Region	479..719	/note="region of homology to ATPase/helicase family SWI2/SNF proteins"
FT	Misc-difference	705	/note="Iys in ecotype Columbia"
FT	Peptide	832..838	/note="nuclear localization motif"
FT	Peptide	858..862	/note="nuclear localization motif"
FT	Domain	995..1015	/note="predicted membrane-spanning domain"
FT	Misc-difference	1219	/note="asp in ecotype Columbia"
FT	Region	1462..1672	/note="internal repeat region"
FT	Region	1848..1894	/note="internal repeat region"
FT	Domain	1899..1941	/note="actin-binding domain"
XX	WO200100801-A2.		
XX	04-JAN-2001.		
XX	21-JUN-2000; 2000WO-EP05761.		
XX	23-JUN-1999; 99GB-0014623.		
XX	(NOVS) NOVARTIS AG.		
XX	(NOVS) NOVARTIS-BREITNUNGEN VERW GES MBH.		
XX	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;		
XX	WPI: 2001-137952/14.		
XX	N-PSDB: AAA89353, AAA89354.		
XX	Novel gene encoding a protein that controls gene silencing, in particular silencing of plant genes -		
XX	Claim 1: Page 32-39; 48pp; English.		
XX	The present sequence is that of Arabidopsis thaliana protein MOM, which controls gene silencing, particularly gene silencing. It is encoded by a novel silencer gene (see AAA89353) of A. thaliana ecotype zurich. The protein exhibits partial similarity with ATPase/helicase proteins of the SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219 compared with the protein (see AAA20063) encoded by the silencer gene of wild-type A. thaliana ecotype Columbia. Gene silencing is useful as a molecular tool for regulating gene expression.		
XX	Sequence 2001 AA:		
XX	Query Match 20.1%; Score 105.5; DB 22; Length 2001; Best Local Similarity 6.4%; Pred. No. 1.5;		
XX	Matches 82; Conservative 56; Mismatches 128; Indels 135; Gaps		
QY	2 GILPSCSTMGVYRVYRDDHVENSTKSGVHDLGNDYVCGSPYITIRMAAE 61		
DB	1645 gnhpdavniigldntsvverhpsig-sdca-----emeisepgyverstfanllhegvgh 17000		
QY	62 VLEGYTVYVCARPMAPLEROKRVVHRSKIL-----PFEKRGVYRELRSK 111		
DB	1701 --sagvtalv-----psllnngtqlavqvpqipfvpfnpflth-elek1 1743		
QY	112 THRSQSSKRYF-DYHSVTMOOLEIRINRGEGPEHLGDRPSKHUKLFTIR--HCLRA---L 165		

D	b1744	rresenbkktfeekksllkxaele-----rkmaevqgefirkfheesehnt	1769
Oy	166	RLEPRLSID-----LMESPPLNLSGALSPRTAKDELITOMLKSAAENSELGMYXK	216
D	1790	rltklekcdhlvmxklilanaflskctckdkkspsgaprykldlqiaqraysalrnyiap	1849
Oy	217	ROEFLYLRARRRRRKFRAMKEVLIQISISMKPWEFFHTPMAYRDSGSP-----p	262
D	1850	qr-----lgsstfpapalvsapiqlqssfdpapypapljapqasssf	1890
Oy	263	KNAATPS-----LPGEK-----NISPRQVS-----VPORSSPPPKNVSP	297
D	1891	savsrpsalllnfavcnpqpgrqlslsnlaplpasvtpatnpgljiraphahinsyrpsstfp	1950
Oy	298	-----PPQPAFVARTSKYSKASGOOVORNRGNAXSL	328
D	1951	vataprtsvvpq-----altysavslqqgqeqqgqgal	1984
RESULT 4			
ID	AAB20063	AAB20063 standard; Protein: 2001 AA.	
AC	AAB20063;		
XX			
DT	23-APR-2001	(first entry)	
XX			
DE	Arabidopsis thaliana silencing gene-encoded protein.		
XX			
KW	Gene silencing; silencing gene; MOM.		
XX			
OS	Arabidopsis thaliana.		
XX			
FH	Key	Location/Qualifiers	
FT	Region	177..350	
FT	Peptide	/note= "internal repeat region"	
FT	Binding-site	362..367	
FT	Region	/note= "nuclear localization motif"	
FT		460..467	
FT		/note= ATP/GTP-binding motif"	
FT		479..719	
FT		/note= "region of homology to ATPase/helicase family SWI2/SNF proteins"	
FT	Misc-difference	705	
FT	Peptide	/note= "Met in ecotype zürich"	
FT		832..838	
FT	Peptide	/note= "nuclear localization motif"	
FT		858..862	
FT	Domain	/note= "nuclear localization motif"	
FT		995..1015	
FT	Misc-difference	1219	
FT	Region	/note= "predicted membrane-spanning domain"	
FT		1462..1672	
FT	Region	/note= "internal repeat region"	
FT		1848..1894	
FT	Domain	/note= "internal repeat region"	
FT		1899..1941	
FT		/note= "actin-binding domain"	
PD	MO200100801-A2.		
XX			
PD	04-JAN-2001.		
XX			
PF	21-JUN-2000;	2000MO-EPOS761.	
XX			
PR	23-JUN-1999;	99GB-0014623.	
XX			
XX	(NOVS) NOVARTIS AG.		
PA	(NOVS) NOVARTIS-ERFINDUNGEN VERN GES MBH.		
PI	Habu Y, Miteilsten Scheid O, Amedeo P, Paszkowski J;		

```
XX WP1: 2001-137952/14.
DR N-PSDB: AAA89385.
XX
PT Novel gene encoding a protein that controls gene silencing, in
PT particular silencing of plant genes
PS
PS Claim 8: Page -: 48pp: English.
XX
CC The present sequence is that of Arabidopsis thaliana protein WOM,
CC which controls gene silencing, particularly gene silencing. It
CC is encoded by a novel silencer gene (see AAA89385) of A. thaliana
CC ecotype Columbia. The protein exhibits partial similarity with
CC ATPase/helicase proteins of the SWI2/SNF2 family. The sequence
CC differs at amino acids 705 and 1219 compared with the protein
CC (see AAB20062) encoded by the silencer gene of Arabidopsis thaliana
CC ecotype Zurich. Gene silencing is useful as a molecular tool for
CC regulating gene expression.
CC Note: the present sequence is not shown in the specification but
CC is derived from the A. thaliana mutant silencer gene protein
CC sequence given in AAB20062.
XX
XX Sequence 2001 AA:
SQ
Query Match 6.1%; Score 105.5; DB 22; Length 2001;
Best Local Similarity 20.4%; Pred. No. 1.5; Mismatches 128; Indels 135; Gaps 19;
Matches 82; Conservative 56;
OY 2 GLPFCSTVVGCTVAVVPRDGHVENLSKCSVHDLILGNPRYYCGSTPYTTINMAAE 61
Db 1645 gnhpdtavtldtctvaephlsq-sdac---emeisepqyqvsrftanlthegveh 1700
OY 62 VLEGYTVFVCATPNAQPFLEKRPKVHRSKIL-----PRFSKHGVHRELASP 111
Db 1701 --sagvtalv-----psllnngteqlavqvpqdpfpvfnqpfh--elekl 1743
OY 112 THSGOGRKRVF-DYHSTVMQQLSEIRNEGPEHLADGPSKHLKLVFIR--HCLRA--L 165
Db 1744 rresenakktfeekskllkaele-----lkmeevgaefrrtfheveeinhnt 1769
OY 166 RLPRISID-----LMESPLPNLSGEMLSPTATADETTQMLKSAKSLGVYK 216
Db 1790 rtklekdknlvilmkllanaflskcdkavspsqagpqrklqqlaqaqsalrnylap 1849
OY 217 ROEFYLRARRRRRRKFMKRVQISISEMKPVMEFHTPMAYRDSGSP-----P 262
Db 1850 qg-----lqasfapal-vsaqlqlqgsfipagpapqlpqasfip 1890
OY 263 KMSSTPS-----LSPK-----NISPPKQVS-----VPSRSSPPKVNVP 297
Db 1891 ssvsrrpalllnfaycpcmpqprqpllnlptprvtpatpnlglspaphlnsyfpassrp 1950
OY 298 -----PQPAFVARTASKYSAAQOVQRNRGNKAKSL 328
Db 1951 vatatpcsvvpq-----alysavslqgqgqgqpqgsll 1984
RESULT 5
AAU22961
ID AAU22961 standard; Protein: 560 AA.
XX
AC AAU22961:
XX
XX 18-DEC-2001 (first entry)
DE
XX
XX Novel human enzyme polypeptide #47.
XX
KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW lysase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
```

```
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0224519.
XX
XX 14-AUG-2000; 2000US-0225213.
XX
XX 14-AUG-2000; 2000US-0225214.
XX
XX 14-AUG-2000; 2000US-0225266.
XX
XX 14-AUG-2000; 2000US-0225267.
XX
XX 14-AUG-2000; 2000US-0225268.
XX
XX 14-AUG-2000; 2000US-0225270.
XX
XX 14-AUG-2000; 2000US-0225447.
XX
XX 14-AUG-2000; 2000US-0225757.
XX
XX 14-AUG-2000; 2000US-0225758.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 18-AUG-2000; 2000US-0226279.
XX
XX 22-AUG-2000; 2000US-0226681.
XX
XX 22-AUG-2000; 2000US-0226686.
XX
XX 22-AUG-2000; 2000US-0227182.
XX
XX 23-AUG-2000; 2000US-0227209.
XX
XX 30-AUG-2000; 2000US-0228324.
XX
XX 01-SEP-2000; 2000US-0228287.
XX
XX 01-SEP-2000; 2000US-0229343.
XX
XX 01-SEP-2000; 2000US-0229344.
XX
XX 01-SEP-2000; 2000US-0229345.
XX
XX 05-SEP-2000; 2000US-0229509.
XX
XX 05-SEP-2000; 2000US-0229513.
XX
XX 06-SEP-2000; 2000US-0230437.
XX
XX 06-SEP-2000; 2000US-0230438.
XX
XX 08-SEP-2000; 2000US-0231242.
XX
XX 08-SEP-2000; 2000US-0231243.
XX
XX 08-SEP-2000; 2000US-0231244.
XX
XX 08-SEP-2000; 2000US-0231245.
XX
XX 08-SEP-2000; 2000US-0231413.
XX
XX 08-SEP-2000; 2000US-0231414.
XX
XX 08-SEP-2000; 2000US-0231415.
XX
XX 08-SEP-2000; 2000US-0232080.
XX
XX 08-SEP-2000; 2000US-0232081.
XX
XX 12-SEP-2000; 2000US-0231968.
XX
XX 14-SEP-2000; 2000US-0232397.
XX
XX 14-SEP-2000; 2000US-0232398.
XX
XX 14-SEP-2000; 2000US-0232399.
XX
XX 14-SEP-2000; 2000US-0232400.
XX
XX 14-SEP-2000; 2000US-0232401.
XX
XX 14-SEP-2000; 2000US-0233063.
XX
XX 14-SEP-2000; 2000US-0233064.
XX
XX 14-SEP-2000; 2000US-0233065.
XX
XX 21-SEP-2000; 2000US-0234223.
XX
XX 21-SEP-2000; 2000US-0234274.
XX
XX 25-SEP-2000; 2000US-0234997.
```


KM Human; extracellular matrix and cell adhesion molecule; XMAP;
 KM gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KM Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KM sickle cell anaemia; thalassemia; autoimmune disorder; adenocarcinoma;
 KM inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KM Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KM Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KM osteoarthritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KM osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KM infection; cell proliferative disorder; actinic keratosis; myeloma;
 KM arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotoxic;
 KM neuroprotective; dermatological.

XX Homo sapiens.
 XX OS
 XX Key Location/Qualifiers
 XX FH 20..211
 XX FT Region /note="Leucine rich repeat"

XX NO200142285-A2.

XX 14-JUN-2001.

XX 05-DEC-2000; 2000MO-US32990.

XX 10-DEC-1999; 99US-0172852.

XX 16-DEC-1999; 99US-0172354.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Lai P, Burford N, Azimzal Y, Patterson C;
 PI Baughn MR, Lu DM, Shah P, Au-Young J;
 DR N-PSDB; AAD08051.

XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis,
 PT prevention, treatment of genetic, autoimmune and cell proliferative
 PT disorders -

XX Claim 1; Page 100-101; 135pp; English.

XX The present sequence is a human extracellular matrix and cell
 CC adhesion molecule (XMAP). The XMAP is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAP. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XMAP. The
 CC polynucleotides encoding XMAP are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XMAP. They are also used for generating
 CC hybridisation probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XMAP may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XMAP, or in assays to monitor patients being treated with XMAP. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.

SO Sequence 377 AA;

Query Match 6.0%; Score 104.5; DB 22; Length 377;
 Best Local Similarity 19.3%; Pred. No. 0.19;
 Matches 52; Conservative 31; Mismatches 101; Indels 85; Gaps 9;

QY 98 FSKGVHVRREASPTRGSGQSRKVD--YHSVTMOQLSIRN----- 137
 DB 19 ftegrredhelirnrhbgpskshelpyglahqpfkryskclhrckrpkpp 78
 QY 138 -----EGPEPLADRPSSKHLKVFTRCRLRLRPISIDLMESLPALSEALSPARA 193
 DB 79 nkppkfnpn-----qppkh----- 107
 QY 194 KDEITOMILKSAARS-----ELGMVYSKROEFLRRARRRRKFAWKPYLQISIE 242
 DB 108 rtqpsvctfpsaactklitipnvtfllpqnatlisrenvntssavlatlpvnpapqdtla 167
 QY 243 MKPWMEFHTPMAYRDSGSPKNAFPLPG-----PKNISPPROVS-----VPO 286
 DB 168 appcpstatp-appssappettaappcpstatqppssappettaappcpstatpapp 226
 QY 287 RSPPEPKNVSPPOPAFVARFASKISAAS 315
 DB 227 sssappettaappps--atlpdlssaa 253

RESULT 7

ABBS814
 ID ABBS814 standard; Protein; 1284 AA.

XX AC ABBS814;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 3234.

XX KM Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;
 DR N-PSDB; ABL02917.

XX WPI: 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Disclosure: SEQ ID NO 3234; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 1284 AA:

Query Match 6.0%; Score 104; DB 22; Length 1284;
 Best Local Similarity 20.6%; Pred. No. 1.1; Mismatches 19;
 Matches 79; Conservative 43; Indels 154; Gaps 19;

OY 16 RVYFPGHVENLSKCSVH-----DLLGNPDYV-----CGSTPYTINRM 57
 Db 631 imlfrqgahmsrakgmilnatwgrllkkrnphlvhngsagsstlgspsmuvygsr 690
 OY 58 AAEELEYGVTFVCAFPNAOPELRQPKVYHRSKILRFRKHHVHRELRSFTHSQO 117
 Db 691 dseep-----lartpsdallvepepy-----tpqeqa 717
 OY 118 SKKVPDHSVTMQQLESIRNEGPEP--HLAGDRP-SKHKLKLVTRHCLRALRLPRISID 173
 Db 718 qllefdpdaqinehvet-pgeypdpdaasglsgmtplsmhmq-----gisv1 762
 OY 174 LMESP-----LPNLS-----GEALSPTA-----TAKDEITOMILKS 204
 Db 763 ppeppvatgaagrpnlslqmpgaskyqylggrrtaapltappppvylkasttplldq 822
 OY 205 AARSELG-MVYSKROEFLRRARRRRKFAKPKVLOSISEMKFVMEFHTPMAYRDSGSPK 263
 Db 822 evlptkikllyvssgqyvaqgs-----spilq-----epalptlygnsa--- 862
 OY 264 NASTPSPLEGGKNISPPROYSVORSSPPPKNVSPPOPAF-----VARTASKYSA 313
 Db 863 -agapfqipq-----aqrqekqp-----pqgqplyanqylnvakaaaaavyap 906
 OY 314 AS-----QGVORNRGNAKSL 328
 Db 907 ssstlmsnqgqgqgrinvargl 929

RESULT 8

AAR15160 AAR15160 standard; Protein; 750 AA.

XX AAR15160;

AC 17-DEC-2001 (updated)
 DT 28-FEB-1992 (first entry)

DE FB29 chicken c-ski protein.

KW Transgenic; muscle growth; muscle degeneration; obesity.

OS Gallus domesticus.

PN USN7620415-N.

PD 05-NOV-1991.

PF 03-DEC-1990; 900S-0620415.

PR 03-DEC-1990; 900S-0620415.

PA (USSH) NAT INST OF HEALTH.

PI Hughes S, Suttrave P;

DR MPI: 1991-361684/49.

DR N-PSDB: AA014799.

PT DNA encoding chicken C-ski proteins - used to produce transgenic
 PT animals having increased muscle growth, and to treat muscle
 PT degeneration or obesity

XX Disclosure; Fig 2; 55pp; English.
 PS
 XX

CC This is the amino acid sequence of the chicken c-ski protein FB29.
 CC It may be produced in transgenic animals, using a DNA construct, to
 CC stimulate muscle growth or prevent muscle degeneration. It can be
 CC used to produce domestic livestock with increased muscle size and
 CC decreased fat tissue. It may also be used to treat serious muscle
 CC injury or muscle degenerative diseases, or patients suffering from
 CC obesity.

CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NRTS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpl/updates/nrtis_us.html.)

SQ Sequence 750 AA:

Query Match 6.0%; Score 103.5; DB 12; Length 750;
 Best Local Similarity 21.0%; Pred. No. 0.6;
 Matches 81; Conservative 40; Mismatches 134; Indels 131; Gaps 17;

OY 1 MGLPRSCYTVGQYRVYRPGHVENLSKCSVHDLGLGNPDYVCGSTPYTINRMAE 60
 Db 147 mgllptisaps-----cglltklthaerlc-- 169
 OY 61 EYLEYGVTFVCAFPNAO-----PFLERQPKVYHRSKILRPSK-HGVHRELRS 110
 Db 170 nallylgty----pnhckkqfstleleltekstfkynes-----fgckqdlvpeyls 219
 OY 111 -PTHGSOQ---SRKVPDHSVTMQQLESIRNE-----GPE--- 141
 Db 220 nptlaactqlchcrimpphkfvnshkslqnstcmwgfdsanwtsy1llsqdytqkeka 279
 OY 142 --PFLAGDRPSK-HLKLVTFRHCLRALRUPRISI-----DLMESPLPNL 182
 Db 280 tlpqllidemkekfdynkkyrkpprntresrvqylrpnkmfktm1wopaagsav1lqpda 339
 OY 183 SGELSPYATAKDEITOMILKSARSELGMVYSKROEFLRRAR-----RRK 230
 Db 340 neypsdpaskkktldpsqspaste-----kekqswlslssasnskysicvprqrl 394
 OY 231 FAKPKVLOSISEMKFVMEFHTPMAYRDSG---SPPKNASTP--SLRGP---KNISPPRO 281
 Db 395 saifpwpavasaanekeisthlpallirdssfyskysfenavaav1alppaqkvvsnpcc 454
 OY 282 VSVFORSSPPPKNVSPPOPAFYART 307
 Db 455 atlvpspseppsaaqpkkhaaet 480

RESULT 9

AAR96994 AAR96994 standard; Protein; 1321 AA.

XX AAR96994;

AC 16-AUG-1996 (first entry)

DE Mouse IRS-2.

KW IRS-2; insulin receptor substrate-2; diabetes; therapy;
 KW diagnosis; vector; antibody; transgenic animal.

OS Mus sp.

XX

XX

XX

XX

XX

XX

Location/Qualifiers
 30..141
 /label= IH1
 /note= "IRS-homology domain I"
 190..366
 /label= IH2
 /note= "IRS-homology domain II"

FT	Domain	481..527	
FT		/label= I13	
FT	Modified-site	/note= "IRS-homology domain III"	
FT		74..80	
FT	Modified-site	/note= "common phosphorylation site"	
FT		120..125	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		135..140	
FT	Modified-site	/note= "common phosphorylation site"	
FT		180..185	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		499..504	
FT	Modified-site	/note= "common phosphorylation site"	
FT		535..540	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		537..542	
FT	Modified-site	/note= "common phosphorylation site"	
FT		593..598	
FT	Modified-site	/note= "common phosphorylation site"	
FT		623..628	
FT	Modified-site	/note= "common phosphorylation site"	
FT		627..632	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		648..653	
FT	Modified-site	/note= "common phosphorylation site"	
FT		670..675	
FT	Modified-site	/note= "common phosphorylation site"	
FT		733..738	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		757..762	
FT	Modified-site	/note= "common phosphorylation site"	
FT		813..818	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		899..903	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		910..915	
FT	Modified-site	/note= "common phosphorylation site"	
FT		969..974	
FT	Modified-site	/note= "common phosphorylation site"	
FT		1031..1036	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		1060..1065	
FT	Modified-site	/note= "unique phosphorylation site"	
FT		1241..1246	
FT	Modified-site	/note= "common phosphorylation site"	
FT		1302..1307	
FT	Modified-site	/note= "common phosphorylation site"	
XX			
PN	W09610629-A1.		
PD			
XX	11-APR-1996.		
XX			
XX	03-OCT-1995;	95WO-US13041.	
XX			
XX	03-OCT-1994;	94US-0317310.	
XX			
PA	(JOSL-) JOSLIN DIABETES CENT INC.		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Pierce JH, Sun XJ, White MF;		
XX			
XX	WPI; 1996-209351/21.		
DR	N-PSDB; AAT28293.		
XX			
PT	New Insulin receptor substrate polypeptide and corresp. nucleic acid		
PT	- vectors, antibodies etc., useful for diagnosis, treatment and		
PT	assessing risk of diabetes etc., also for drug screening		
XX			
PS	Disclosure; Page 47-53; 105pp; English.		
CC	Mouse Insulin receptor substrate-2, or IRS-2 (AAR06994), is a		
CC	substrate for the Insulin receptor, Interleukin-4 receptor and		
CC	Interleukin-15 receptor, and can be phosphorylated by these		

	C	c	receptor". It contains a number of phosphorylation sites in
	C	C	common with IRS-1, showing the 2 proteins to be related.
	C	C	can be obd., e.g., by affinity purification from insulin-stimulated
	C	C	FDC-P2 cells using immobilized shz domains of p85, or can
	C	C	be expressed in transfected host cells carrying an encoding cDNA
	C	C	sequence (ARF8293). Cells or animals having the IRS-2 transgene
	C	C	can be used to study insulin-related disorders, e.g., type II
	C	C	dibetes. IRS-2 can be used to raise antibodies and for drug
	C	C	screening.
SQ	X	x	Sequence 1321 AA;
Query Match			6.0%; Score 103.5; DB 17; Length 1321;
Best Local Similarity			22.2% ; Pred. No. 1.3;
Matches 72;			Conservative 30; Mismatches 93; Indels 133; Gaps 16
D	b	B	
Y	y	Y	30 SCGVHDLGNDGYVCGSPPIYTIRNMAAEVLVEYGTVFCATPNAPFLERGPKNVH 89 : : : : : : : : : D b s c s -----gdnqyvlmspp-----vgrlleeeite-----pqatpgagltf----- 841
Y	y	Y	90 RGSKILRFRESKHCHVAEELNSPTHSQSQRKYFDIHSVTWQLSETRNECPPEHLADGRP 149 : : : : : : : : : D b 842 -----gaagqshtlc-----phhsavps-----smr-----psal99rp 869
Y	y	Y	150 SKHLKFIVFHICRLRLPRISID-----LMESPPLNLSCBALSPATAKDEITOMTL 202 : : : : : : : : : D b 870 eg-----flgcrcravrlptllslglqltlpmgeplrt---epksp----- 908
Y	y	Y	203 KSAARELGNMYRKQEFLLRARRRRRAKKMPYLGISEMKRVMEPHTPMAYRDSGPSPP 262 : : : : : : : : : D b 909 -----geyin-----ldfgsegtlisppepllaaaassssllaaspasssiagslrpg 956
Y	y	Y	263 KWA-----STPSLEPGPKN-----ISPFRQVSVPQAASSPPRNK 294 :: :: - - - - - - - - - D b 957 tsdsdtqrplsldymnlafsfskpkprstcrsgdlvgmdglisr--easspylpplrps 1014
Y	y	Y	295 VSP-----PPQPAFVARTKSKYSASQ 316 : : : : : : D b 1015 tepselqgpilppeagdlyylrlpasaltsq 1043
RESULT 10	AAB23012		
ID	AAB23012 standard; Protein:	2742 AA.	
XX	AA23012;		
XX	16-JAN-2001 (first entry)		
DE	Human APC protein (splice variant 2).		
KW	APC gene; Adenomatous polyposis Coli gene; human; chromosome 5q21;		
KW	familial adenomatous polyposis; RAP locus; Gardner's syndrome; GS;		
KW	sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;		
KW	bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;		
KW	tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;		
KX	genetic predisposition; drug screening; DPZ.5; splice variant..		
OS	Homo sapiens.		
PN	US611412A-A.		
PD	05-SEP-2000.		
PF	25-MAY-1995;	95U5-04505B2.	
PR	16-JAN-1991;	91GB-000096Z.	
PR	16-JUN-1991;	91GB-000096J.	
PR	16-JAN-1991;	91GB-000097A.	
PR	08-AUG-1991;	91UB-000097S.	
PR	12-AUG-1994;	91US-074194O. 94US-028954B.	

KM screening; prostate cancer; benign prostatic hyperplasia;
 KW hair loss; alopecia; therapy.

XX Homo sapiens.

OS WO200172332-A1.

XX 04-OCT-2001.

XX 26-MAR-2001; 2001WO-US09378.

XX 24-MAR-2000; 2000US-191768P.

XX 15-AUG-2000; 2000US-225618P.

XX (UNIV) UNIV NEW YORK STATE.

XX Garabedian M, Taneja S, Hittelman A, Markus S;

DR WPI; 2001-611582/70.

XX N-PSDB; AAH26896.

PT New androgen receptor transcriptional coregulatory proteins that
 PT interact with androgen receptor to regulate androgen-dependent gene
 PT expression, useful for producing antibodies to treat androgen dependent
 PT diseases

PS Claim 31; Page 99-107; 113pp; English.

XX The present sequence is that of a novel human androgen receptor
 CC trapped protein (AR), which belongs to a new class of proteins
 CC that interact with the N-terminus of the androgen receptor as
 CC androgen receptor transcriptional coregulatory (i.e. coactivator)
 CC proteins, regulating androgen-dependent gene expression. To identify
 CC proteins that interact with the androgen receptor N-terminus,
 CC a modified yeast two-hybrid system that allowed the identification
 CC of factors expressed in the prostate and which associated with
 CC transcriptional activators was used. An androgen-stimulated LNCaP
 CC prostate cancer cell cDNA library fused to the Lexa DNA binding
 CC domain was screened for proteins that interacted with the androgen
 CC receptor N-terminal transcriptional activation domain (amino acid
 CC residues 18-500). From approximately 1 million library
 CC transformants, 8 clones were isolated, and termed ARts, for
 CC androgen receptor trapped. ARts are useful for producing
 CC antibodies (or molecules having the binding portion of an antibody)
 CC which can be administered to patients to treat androgen dependent
 CC diseases, especially prostate cancer, benign prostatic hyperplasia
 CC or androgen-dependent hair loss. They can be used to identify
 CC inhibitors disrupting interaction between an androgen receptor and
 CC an androgen receptor transcriptional coregulatory protein, useful
 CC to inhibit such interaction (all claimed) therapeutically.

XX Sequence 2783 AA;

Query Match 5.9%; Score 103; DB 22; Length 2783;
 Best Local Similarity 28.0%; Pred. No. 4.1;

Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;

OY 171 SIDLME-----SPLDMLSGEALSPFA--TAKDELITOMILKSAARSELGMVYSKQ 218

DB 1454 smdamelilptsscsctmpsqaysapapamntassafldl---taaeelatlnskke 1510

OY 219 EFYLRARRRRRRKFAV-----KPVLOSISEMKPVMEFHTPMAYRDSGSP 262

DB 1511 -----agdekpklaeapsaqpnqtqekgqpkpeldq--gqeqpqdktnrpqg----klp 1558

OY 263 KNASTPSLPGPNISIPRQVSVPRSSPPPKNNSP 299

DB 1559 qlvslpslpgpppgappqpcplpg-sspspsqlshlp 1594

RESULT 14
 AAR63508

ID AAR63508 standard; Protein; 2842 AA.

XX AAR63508;

XX 23-MAY-1995 (first entry)

XX Adenomatous polyposis coli tumour repressor.

XX Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;

XX familial adenomatous polyposis; cancer diagnosis and prognosis;

XX tumorigenesis suppression.

XX Homo sapiens.

XX US5352775-A.

XX 04-OCT-1994.

XX 08-AUG-1991; 91US-0741940.

XX 16-JAN-1991; 91GB-0000962.

XX 16-JAN-1991; 91GB-0000963.

XX 16-JAN-1991; 91GB-0000974.

XX 16-JAN-1991; 91GB-0000975.

XX (CANC-) CANCER INST.

XX (ICIL) IMPERIAL CHEM IND.

XX (UYJO) UNIV. JOHNS HOPKINS.

XX (UTAH) UNIV UTAH.

XX Albetson H, Anand R, Carlson M, Groden J, Hedge RJ,

XX Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;

XX Vogelstein B, White RJ;

XX WPI; 1994-316233/39.

XX N-PSDB; AAQ72297.

XX Claim 3; Columns 71-84; 113pp; English.

XX AAQ72297 is a cDNA isolated from the human adenomatous polyposis
 CC coli (APC) gene. It encodes the tumour repressors described in AAR63507
 CC and AAR63508. Determination of alterations in APC or its expression
 CC products, can be used for the diagnosis and prognosis of cancer e.g.
 CC colorectal, lung and breast tumours; and for determining predisposition
 CC to certain cancers such as familial adenomatous polyposis (FAP) and
 CC Gardner's syndrome. The wild type APC gene (or a part of it) can be
 CC used therapeutically to restore gene function, while primers and probes
 CC derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to
 CC detect mutations. Also APC proteins or analogues can be administered to
 CC animals carrying a mutated APC allele are useful for detecting
 CC therapeutic agents able to suppress tumorigenesis.

XX Sequence 2842 AA;

Query Match 5.9%; Score 103; DB 15; Length 2842;
 Best Local Similarity 22.3%; Pred. No. 4.2;

Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

OY 79 PLELRQPKVIRVRSKILPRFSKGVNRELSPTHSQOSRKVFEDYHSVTMOQLESTRNE 138

DB 2151 pqdeekspftmskprllkpgekstletkleskglkgkv--ykslllgkvrs---- 2204

OY 139 GPEPHIADDRPSKHLKIVFIRHCLRALRIPRISIDLMESPLML--SGEALSPATAKDE 196

DB 2205 --nseisgq-----mkqplga-nmpslsrgtmtlpgvynssasstspvskkqpp 2251

OY 197 ITOMILKSAARSELGMVYSKROEFLIRARRRRRRKFAKMPYLOSISEMKPVMEFHTPMAYR 256

DB 2252 Iktpsakapsqglat-tspr-----gakpsvk--selspvar-qtsq19g 2293
 QY 257 DSGSPPKNA---STPSLPGPKNISPPQVSVPORSSPPPKNVSPPOPAFVARTASKYSA 313
 DB 2294 skkapsrsgsrdsfprpdqqlsrpqlspgpnslspgnglsppnklsq1prtspssta 2353
 QY 314 ASQOVQRNRGNKSLY 329
 DB 2354 st-----kssgsgkmsy 2365

RESULT 15

AAR26052 AAR26052 standard; Protein; 2843 AA.

AC AAR26052;

DT 28-JAN-1993 (first entry)

DE APC gene product in familial adenomatous polyposis.

KW neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;
 KW prognosis; treatment; sporadic colorectal carcinomas; ss.

OS Homo sapiens.

XX MO9213103-A.

XX 06-AUG-1992.

XX 16-JAN-1992; 92MO-US00376.

XX 16-JAN-1991; 91GB-0000963.

XX 08-AUG-1991; 91US-0741940.

XX (CANC-) CANCER INST.

XX (ICIL) IMPERIAL CHEM IND PLC.

XX (UYJO) UNIV JOHNS HOPKINS.

XX (UTAH) UNIV UTAH.

XX Albertsen H, Anand R, Carlson ML, Groden JL, Hedge PJ;
 PI Joshiyng K, Kinzler KW, Markham A, Nakamura Y, Thliveris A;
 PI Vogelstein B, Whiteerl, Markham AF;

XX WPI; 1992-284685/34.

XX N-PSDB; AAO27234.

XX Detection of somatic and germ-line alterations of human APC gene
 PT - used to diagnose, treat and study familial adenomatosis
 PT polyposis and sporadic colorectal cancer

XX Disclosure: Page 47; 132pp; English.

XX This sequence is encoded by the APC (Adenomatous Polyposis Coll)
 CC gene associated with tumorigenesis, found on chromosome 5q.
 CC The sequence may be mutated by deletions, insertions, inversions, or
 CC point mutations of the gene. The APC gene is expressed in most normal
 CC tissues as well suggesting that APC is a tumour suppressor.

XX Sequence 2843 AA;

XX

Query Match 5.9%; Score 103; DB 13; Length 2843;

Best Local Similarity 22.3%; Pred. No. 4.2; Mismatches 107; Indels 46; Gaps 11;

Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEEROPKVVYRSGKILPRFSKGVHRELRSPTGSOQSRKVDHASTVMQOLESTRNE 138

DB 2152 pdgeekrftsmngprllkpgkstletkkeskylg9kvv--ykslltgvksr--- 2205

QY 139 GPEPHLAGDRPSKHLKLVFIRHCLRALRLPRISIDLMESPLNL--SGEALSPTATKDE 196

DB 2206 --nselsqg-----mkqplga-nmpslsrgtmtlhpvyrnssstlprvskk9pp 2252
 QY 197 ITOMILKSAARSELGMYSKRQEFYLRARRRRKFAWKPEVLOSISEKKPYMEFHTPMAYR 256
 DB 2253 Iktpsakpsqglat-tspr-----gakpsvk--selspvar-qtsq19g 2294
 QY 257 DSGSPPKNA---STPSLPGPKNISPPQVSVPORSSPPPKNVSPPOPAFVARTASKYSA 313
 DB 2295 skkapsrsgsrdsfprpdqqlsrpqlspgpnslspgnglsppnklsq1prtspssta 2354
 QY 314 ASQOVQRNRGNKSLY 329
 DB 2355 st-----kssgsgkmsy 2366

Search completed: September 16, 2002, 00:42:28
 Job time: 3377 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2002, 23:57:51 ; Search time 24 Seconds
(without alignments)
336.870 Million cell updates/sec

Title: US-09-828-313-35
Perfect score: 1734
Sequence: 1 MGLTFSCVTVGGYRVVYP.....SAASQGVGRNKGNAKSLXMA 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
 - 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
 - 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
 - 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
 - 5: /cgn2_6/ptodata/2/1aa/PCUTS.COMB.pep.*
 - 6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	6.7	912	5 PCT-US95-03747-2	Sequence 2, Appl1
2	103.5	6.0	1321	1 US-08-317-310A-64	Sequence 64, Appl1
3	103	5.9	2842	1 US-07-741-940-7	Sequence 7, Appl1
4	103	5.9	2842	1 US-08-289-548A-7	Sequence 7, Appl1
5	103	5.9	2842	1 US-08-452-654-7	Sequence 7, Appl1
6	103	5.9	2843	1 US-07-741-940-2	Sequence 2, Appl1
7	103	5.9	2843	1 US-08-289-548A-2	Sequence 2, Appl1
8	103	5.9	2843	1 US-08-452-654-2	Sequence 2, Appl1
9	103	5.9	2843	1 US-08-452-655B-2	Sequence 2, Appl1
10	103	5.9	2843	1 US-08-452-655B-2	Sequence 2, Appl1
11	103	5.9	2843	2 US-08-370-235A-2	Sequence 2, Appl1
12	103	5.9	2843	2 US-08-450-582-2	Sequence 2, Appl1
13	103	5.9	2843	2 US-08-450-582-7	Sequence 7, Appl1
14	103	5.9	2973	2 US-08-821-355A-7	Sequence 7, Appl1
15	103	5.9	2973	2 US-09-003-687A-7	Sequence 7, Appl1
16	103	5.9	2973	4 US-09-136-605-7	Sequence 7, Appl1
17	101	5.8	679	4 US-09-056-556-194	Sequence 194, App
18	98	5.8	1130	4 US-09-442-100-4	Sequence 4, Appl1
19	97.5	5.6	1589	3 US-08-755-581-189	Sequence 189, App
20	97	5.6	306	1 US-08-217-337-6	Sequence 6, Appl1
21	97	5.6	677	1 US-08-188-582-13	Sequence 13, Appl1
22	97	5.6	677	1 US-08-646-715-13	Sequence 13, Appl1
23	95	5.5	771	2 US-08-742-755-2	Sequence 2, Appl1
24	94.5	5.4	442	3 US-08-834-306-52	Sequence 52, Appl1
25	94.5	5.4	442	3 US-08-993-674A-52	Sequence 52, Appl1
26	94.5	5.4	787	2 US-08-720-484A-4	Sequence 4, Appl1
27	94.5	5.4	787	4 US-08-953-823A-4	Sequence 4, Appl1

28	94	5.4	666	1 US-08-083-590A-17	Sequence 17, Appl1
29	94	5.4	666	2 US-08-346-128-16	Sequence 36, Appl1
30	94	5.4	666	3 US-08-532-384-17	Sequence 17, Appl1
31	94	5.4	1017	4 US-09-600-776-6	Sequence 6, Appl1
32	93	5.4	360	3 US-08-341-018-58	Sequence 58, Appl1
33	93	5.4	560	4 US-08-470-335-194	Sequence 194, App
34	93	5.4	560	4 US-08-470-339-194	Sequence 194, App
35	92.5	5.3	737	1 US-08-188-582-16	Sequence 16, Appl1
36	92.5	5.3	737	1 US-08-646-715-16	Sequence 16, Appl1
37	92.5	5.3	739	3 US-09-035-648-24	Sequence 24, Appl1
38	92.5	5.3	739	4 US-09-001-951-24	Sequence 24, Appl1
39	92.5	5.3	1298	2 US-08-690-473-2	Sequence 2, Appl1
40	92.5	5.3	1298	4 US-09-259-821A-2	Sequence 2, Appl1
41	92.5	5.3	1298	4 US-08-843-659-2	Sequence 2, Appl1
42	92	5.3	404	4 US-09-232-408A-8	Sequence 2, Appl1
43	91.5	5.3	1464	4 US-08-891-640-2	Sequence 8, Appl1
44	91	5.2	345	1 US-08-843-993-1	Sequence 1, Appl1
45	91	5.2	345	3 US-09-059-520A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
PCT-US95-03747-2
Sequence 2, Application PC/TUS9503747
GENERAL INFORMATION:
APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
TITLE OF INVENTION: Brevicin, A Gliad Cell Proteoglycan
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT-US95/03747
FILING DATE: 27-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Imbra, Richard J.
REGISTRATION NUMBER: 37,643
REFERENCE/DOCKET NUMBER: FP-LJ 1453
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-03747-2

Query Match 6.7%: Score 116; DB 5; Length 912;
Best local similarity 24.2%, Pred. No. 0.0047;

Matches 54; Conservative 26; Mismatches 101; Indels 42; Gaps 8;

OY 140 PEPHLAQRPSKHLKVFTRICLRALRPR-----ISIDLMF-----SPLPULS 183
DB 368 PMSHLASDALE--AIYTVETLEELKLPQEAIVESRGAIYSIPTEIDGGCGSSPDEP 424
OY 184 GEALSPATATADDETTOMILKSAARSELGMYSKRQEFYLRARRRRK-----FA 232
DB 425 AEA--PRTLLFE--TGSIVPPLGSSREGKVLQDEKRYRGEERKEEREEVEEDALMA 401

QY 233 WKPVLOSIEKKPVMEFHTPMAYRDSGSPPKNA-----STPSLPGPKNISPPROVSVPQ 286
Db 482 WPELSLSLDPEAPLPTEPVEESLTQASPPVMAALOPGVSPPYDDEAPRPPFVLGPPT 541
QY 287 RSSPPKN---VSPPPQPAFVARTASKYSAAQQOYORNRGNNAK 326
Db 542 KTLPTPREGNLAASPPSTLVGARETEETGCGPELSGAPRGSE 584

RESULT 2

US-08-317-310A-64
; Sequence 64, Application US/08317310A
; Patent No. 5858701
; GENERAL INFORMATION:
; APPLICANT: WHITE, MORRIS F.
; APPLICANT: SUN, XIAO JIAN
; APPLICANT: PIERCE, JACALYN H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,310A
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1321 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-310A-64

Query Match 6.0%; Score 103.5; DB 2; Length 1321;
Best Local Similarity 22.2%; Pred. No. 0.14; 93; Indels 133; Gaps 16;

Matches 73; Conservative 30; Mismatches 93; Indels 133; Gaps 16;

QY 30 SCSEVDLLGNDPEYVCGSTPYTITRMAAEVLEGVYFVCATPNAQPFLEQPKVYH 89
Db 806 SCS-----GDNDQYVLMSSP---VGRILEEERLE-----PQATGAGTF----- 841
QY 90 RGSKITLPRRSKGVHVELRSPTHGSOOSRKVFEDVHSVTWQOLESTRNGPEPHLAGDRP 149
Db 842 -----GAAGSGHTQ-----PHHSVAPS-----SKR-----PSAIGSRP 869
QY 150 SKLKLKVTIRHCLRALRLPRISID-----LMSEPLFNUSGEALSPTATANDETITOMIL 202
Db 870 EG-----FLGQRCRAVVRTRLSLEGLOTLPSMQEYPLPT--EPKSP----- 908
QY 203 KSAARSELGMYSKROEFYLRARRRRKFAWPKVLOSISEKKPVMEFHTPMAYRDSGSP 262
Db 909 -----GEYIN-----IDFGAGCTRLSPAPRPLLAASSSSLSLASSPASLSGTRG 956
QY 263 KNA-----STPSLPGPKN-----ISPPROVSVPORSSPPKN 294

Db 957 TSSDSRORSPLSDYMNLDSESSPKSPKSTNSGDTVGSMDGLLSP--EASSVPPLPDRPS 1014
QY 295 VSP-----PPQPAFVARTASKYSAAQ 316
Db 1015 TSPSSLOQPLPAPAGDLRLPPASAAATSQ 1043

RESULT 3

US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, YUSUKE
; APPLICANT: MAKHURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-07-741-940-7

Query Match 5.9%; Score 103; DB 1; Length 2842;
Best Local Similarity 22.3%; Pred. No. 0.49;

Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEKOPKVVYHRSKILPRFSKHGVHVELRSPTHGSOOSRKVFEDVHSVTWQOLESTRNE 138
Db 2151 PQOEKPTSNKQPRILKPGKSTLETKTIESSEKIGKGVK--YKSLINGKVR----- 2204
QY 139 GEPHLAGDRPSKHLKLVIRHCLRALRLPRISIDLMESPLPNTL--SGEALSPTATANDOE 196

Db 2205 --NSELGQ-----MKOPLQA-NMPSISGRTHIHIPGVNSSSTSPVSKKGP 2251
QY 197 ITOMILKSAASELGMVSKROEFLRRARRRRKFAKPVLOSISEMKPVMEFHTPMAYR 256
Db 2252 LKTPASKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSOIG 2293
QY 257 DSGSPKNA---STPSLPGPKNISPPROVSVPORSSPPKNSVPPQAPAFARTASKYSA 313
Db 2294 SSKAPSRSGSDSTPSRPAQPLSRPIOSPGRNSISPGRNGISPPNKLSQLPRTSSPSTA 2353
QY 314 ASQOVORNRGNKASLY 329
Db 2354 ST---KSSGSGKMSY 2365

RESULT 4
US-08-289-548A-7
Sequence 7, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Homo sapiens
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 5.9%; Score 103; DB 1; Length 2842;
Best Local Similarity 22.3%; Pred. No. 0.49;
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEERPKVYHRSKILPRFSKHGVNRELAPTHSGQSKRVYDHYHVTMOOLESTRNE 138
Db 2151 PDEEKPTTSKNGRILKPEKSTLETKIESKGIKGGKV--YKSLTGKVR----- 2204
QY 139 GEPHLADGPRSKHLKVFTRHCLALRLPRI SIDLMESPDLN--SGEALSPATAKDE 196
Db 2205 --NSELGQ-----MKOPLQA-NMPSISGRTHIHIPGVNSSSTSPVSKKGP 2251
QY 197 ITOMILKSAASELGMVSKROEFLRRARRRRKFAKPVLOSISEMKPVMEFHTPMAYR 256
Db 2252 LKTPASKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSOIG 2293
QY 257 DSGSPKNA---STPSLPGPKNISPPROVSVPORSSPPKNSVPPQAPAFARTASKYSA 313
Db 2294 SSKAPSRSGSDSTPSRPAQPLSRPIOSPGRNSISPGRNGISPPNKLSQLPRTSSPSTA 2353
QY 314 ASQOVORNRGNKASLY 329
Db 2354 ST---KSSGSGKMSY 2365

RESULT 5
US-08-452-654-7
Sequence 7, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 5.9%; Score 103; DB 1; Length 2842;
Best Local Similarity 22.3%; Pred. No. 0.49;
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

OY 79 PLEKOPKVVHRSKILPREFSHGVHVELSPHSGOOSRRVDPYHSVTMOQLESIRNE 138
DB 2151 PDEEKPPTSKNGKRIILKPEKSTLETETKIESEKIGKGV--YKSLITGVKRS--- 2204
OY 139 GPEPLAGDPRSKHLKLVETIRICLRALRPRIISIDLMESPLNL--SGEALSPATAKDE 196
DB 2205 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIGVRNSSSTSPVSKKGP 2251
OY 197 ITOMILKSARSELGMYSKQOEFLRRARRRRKFAKRPVLOSISEKPYMEFHTPMAYR 256
DB 2252 LKTPASKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGC 2293
OY 257 DSGSPKNA---STPSLPQPKNISPPQVSPQSSPPPKNVSPPOPAFVARTASKYA 313
DB 2294 SSNAVSRSRGSDDSTPSRPAQPLSRPIQSPGRNISPGRNGISPPNKLSQLPRTSSPSTA 2353
OY 314 ASQOVRNRGNKAKSLY 329
DB 2354 ST---KSSGSGKMSY 2365

RESULT 6
US-07-741-940-2
Sequence 2, Application US/07741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOF.
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Bitch, Mckie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107, 035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match 5.9%; Score 103; DB 1; Length 2843;
Best Local Similarity 22.3%; Pred. No. 0.49;
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

OY 79 PLEKOPKVVHRSKILPREFSHGVHVELSPHSGOOSRRVDPYHSVTMOQLESIRNE 138
DB 2152 PDEEKPPTSKNGKRIILKPEKSTLETETKIESEKIGKGV--YKSLITGVKRS--- 2205
OY 139 GPEPLAGDPRSKHLKLVETIRICLRALRPRIISIDLMESPLNL--SGEALSPATAKDE 196
DB 2206 --NSEISGQ-----MKOPLQA-NMPSISRGRTMIHIGVRNSSSTSPVSKKGP 2252
OY 197 ITOMILKSARSELGMYSKQOEFLRRARRRRKFAKRPVLOSISEKPYMEFHTPMAYR 256
DB 2253 LKTPASKSPSEGOAT-TSPR-----GAKPSVK--SELSPVAR-QTSQIGC 2294
OY 257 DSGSPKNA---STPSLPQPKNISPPQVSPQSSPPPKNVSPPOPAFVARTASKYA 313
DB 2295 SSNAVSRSRGSDDSTPSRPAQPLSRPIQSPGRNISPGRNGISPPNKLSQLPRTSSPSTA 2354
OY 314 ASQOVRNRGNKAKSLY 329
DB 2355 ST---KSSGSGKMSY 2366

RESULT 7
US-08-289-548A-2
Sequence 2, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOF.
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107, 46943
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 5.9%; Score 103; DB 1; Length 2843;
Best Local Similarity 22.3%; Pred. No. 0.49; Mismatches 107; Indels 46; Gaps 11;
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEROPVYHRSKILPRFSKHGVHRELSPHGSQSRKVFYHSVTMOLESTRNE 138
DB 2152 PDDEKPTSMKGRILKPEKSTLETAKIESKGIKGVK--YKSLITGKVR----- 2205
QY 139 GPEPHLAGDRPSKHLKLVFTRHCLRALRLPRISIDLMESPVLN--SGEALSPYATKDE 196
DB 2206 --NSEISQ-----MKOPLQ--NMPISISGRTHIHGVNRSSTSPVSKKGP 2252
QY 197 ITOMILKSARESELGMVSKROEFYLRARRRRFAWKPVLOSISEMPPVMEFTPMAYR 256
DB 2253 LKTPASKSPSEGTAT--TSPR-----GAKPSVK--SELSVVAR--OTSQIGG 2294
QY 257 DSGSPPKNA---STPSLPGPKNISPPQVSPQRSSPPKNSPPQPAFAVARTASKYSA 313
DB 2295 SSKAPRSRSGSDSTPSRPAQQLSRPIQSPGRNISPGRNGISPPKLSQLPRTISSSTA 2354
QY 314 ASQOVORNRGNKAKSLY 329
DB 2355 ST----KSSGSGKMSY 2366

RESULT 8
US-08-452-654-2
Sequence 2, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 5.9%; Score 103; DB 1; Length 2843;
Best Local Similarity 22.3%; Pred. No. 0.49; Mismatches 107; Indels 46; Gaps 11;
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PLEROPVYHRSKILPRFSKHGVHRELSPHGSQSRKVFYHSVTMOLESTRNE 138
DB 2152 PDDEKPTSMKGRILKPEKSTLETAKIESKGIKGVK--YKSLITGKVR----- 2205
QY 139 GPEPHLAGDRPSKHLKLVFTRHCLRALRLPRISIDLMESPVLN--SGEALSPYATKDE 196
DB 2206 --NSEISQ-----MKOPLQ--NMPISISGRTHIHGVNRSSTSPVSKKGP 2252
QY 197 ITOMILKSARESELGMVSKROEFYLRARRRRFAWKPVLOSISEMPPVMEFTPMAYR 256
DB 2253 LKTPASKSPSEGTAT--TSPR-----GAKPSVK--SELSVVAR--OTSQIGG 2294
QY 257 DSGSPPKNA---STPSLPGPKNISPPQVSPQRSSPPKNSPPQPAFAVARTASKYSA 313
DB 2295 SSKAPRSRSGSDSTPSRPAQQLSRPIQSPGRNISPGRNGISPPKLSQLPRTISSSTA 2354
QY 314 ASQOVORNRGNKAKSLY 329
DB 2355 ST----KSSGSGKMSY 2366

RESULT 9
US-08-452-655B-2
Sequence 2, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Wilcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B

ADDRESSEE: BANNER & WITCOFF, LTD.
STREET: 1001 G STREET, N.W.
City: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/370,235A
APPLICATION DATE: 01-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,48688
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 508 9100
TELEFAX: 202 508 9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-370-235A-2

RESULT 2

E96654

hypothetical protein Fl6P17.12 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E96654

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Hughes, B.; Holzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A66141; MID:21016719

A:Accession: E96654

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: GB:AE005173; NID:948493585; PIDN:AAF75808.1; GSPDB:GN00141

C:Genetics:

A:Gene: Fl6P17.12

A:Map position: 1

Query Match 7.2%; Score 124; DB 2; Length 796;

Best Local Similarity 23.6%; Pred. No. 0.14; Mismatches 132; Indels 52; Gaps 11;

Matches 68; Conservative 36; Mismatches 132; Indels 52; Gaps 11;

74 TPAAC-----PLENQPKVNRGSKILPFGSHGVNRELSPTHSGQSRK 120

Db 164 TPQSQSQSFQHHQSQSPQHNPNGTVSFQNGLP-FQHDHFAQAVQTGSSQDQR 222

QY 121 VPDYNSVTMOLESIRNEGPERPLACD-----RSKHLKLVFTRICLRRLPRISIDLM 175

Db 223 QFPOQQLDEQQAQSMRQOVQGEHSSSGQDMKRPVEEVRILINMATERVSSRRFAEP 282

QY 176 ESDPLRLSGEALSPATAKDEIT-----OMILKSAARSEL-CMYSKROEFLRRARR 228

Db 283 RPSRSFEG---SSHRTPRTELWASKRPVSEPRVRSLELVPMQSEPARQYL--SSRS 337

QY 229 RKFAKMPVLOSISEMKPMVEFHPRMAYRDSG---SPKMASTRPLDPGKNISPPRQVS 284

Db 338 SEAAQLSLPLSVSDS-----HASQPTRSNOSHAVSKRPQYVSKFPHPRPSQPP--TSN 390

QY 285 PQRSSPPKKNVSP-----PQPAFVARTASKYSAASQOVQNR 322

Db 391 PFLPSQPSNSKPFPMSSQSSQNSKPPVSSQSKPLVSSQSK 438

RESULT 3

728666 protein kinase C-related kinase PRKSD - Suberites domuncula

C:Species: Suberites domuncula

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000

C:Accession: T928666

R:Mueller, W.E.C.

submitted to the EMBL data library, April 1997

A:Reference number: Z20501

A:Accession: T928666

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1102 (#NUE>

A:Cross-references: EMBL:Y13104; NID:e1289874; PID:e116786; PIDN:CAA73558.1

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 6.8%; Score 118.5; DB 2; Length 1102;

Best Local Similarity 25.8%; Pred. No. 0.54; Mismatches 149; Indels 59; Gaps 18;

Matches 85; Conservative 36; Mismatches 149; Indels 59; Gaps 18;

Query Match 6.7%; Score 117; DB 2; Length 891;

Best Local Similarity 21.8%; Pred. No. 0.53; Mismatches 137; Indels 92; Gaps 17;

Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

17 VVPDGHEN--LSKSCSVHDLILGNPDY-----VCGSTPYTITNRMAE----- 60

Db 262 VVRDLKRIENILISITGDIKIIDFGLSNLYRQSRRLTFCGSLYFAPELLNAOPYIGPE 321

QY 61 -EVLKGVYVFCATPNMAOPLEBQPKVH---RGSKILPFRSKGVHVELSPTHGS 115

Db 322 VDWMSGIVLYVCGKV-PFDQMSALMAKIKKGVETPS-----LSSDCKGL 371

QY 116 QOSRKYFD-YHSVTMOLES-----TN-EGPEPHLACDRPSKHLKLVFTRICLALMLRP 169

Db 372 LSRMLTDLKRLALEEVLNHPWIMRNEGPPASPAFENSP----- 412

QY 170 ISIDMESPLNLSEALSPATAKDEITOMILKSAARS--ELGMY-----VSKROEFLV 222

Db 413 ITPLDPEIIRMGDFGCPPEKIVRELTKVYISSEAVQSLAKGFGSGNSADKKSFPE 472

QY 223 RRRRRRRKPMKPVLOSISEMKPMVEFHPRM-----AYRSGSPKKNASTP--S 269

Db 473 FRIRHAHDIEPLPLSLSMNTDIYDAFPLISILYLVSERRYVEKGGMMRIKTPVSS 532

QY 270 LRPKNISPPRQVSPQRSSPPKKNVSPQPAFVARTASKYSAASQOVQNRQNAK 326

Db 533 VP-----SSPYV-----PTSYNRTLPPEVY--ATKGDESPKSNVTSLSAR 573

RESULT 4

T40503

protein kinase kin1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000

C:Accession: T40503

R:Gwilliam, R.; Rajadream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.

submitted to the EMBL data library, September 1998

A:Reference number: Z21933

A:Accession: T40503

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-891 (#GWI>

A:Cross-references: EMBL:AL031534; PIDN:CAA20726.1; GSPDB:GN00067; SPDB:SPBC4P6.06

A:Experimental source: strain 972h; cosmid c4P6

C:Genetics:

A:Gene: SPDB:SPBC4P6.06

A:Map position: 2

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

RESULT 5

A38903

protein kinase 1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Sep-1999

C:Accession: A38903; A36474

R:Levin, D.E.; Bishop, J.M.

submitted to GenBank, May 1991

A:Reference number: A38903

A:Accession: A38903

A:Molecule type: DNA

A:Residues: 1-891 <LEV>

A:Cross-references: GB:M64999; NID:g173409; PIDN:AAA63577.1; PID:g173410

R:Levin, D.E.; Bishop, J.M.

Proc. Natl. Acad. Sci. U.S.A. 87, 8272-8276, 1990

A:Title: A putative protein kinase gene (kin1(+)) is important for growth polarity in Sc

A:Reference number: A36474

A:Accession: A36474

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-361, 'R', 363-619, 'T', 621-707, 'W', 709-891 <LE2>

A:Cross-references: GB:M36060

C:Genetics:

A:Gene: kin1+

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase

F:123-395/Domain: protein kinase homology <kin>

F:131-139/Region: protein kinase ATP-binding motif

Query Match 6.7%; Score 117; DB 2; Length 891;

Best Local Similarity 21.8%; Pred. No. 0.53;

Matches 78; Conservative 50; Mismatches 137; Indels 92; Gaps 17;

17 VYRPGHVEN--LSKCSNHDLGPNPY-----VCGSTPTTINRAAE----- 60

Db 262 VVHRRLKIKENILSKTGDKIKIDFELSNLYRQSRKRTGCSGLYFAABELINAPYIGPE 321

QY 61 -EVLEYGVYVCATPNAOPFLERQPKVH-----RSGKILPFSKHGVHRELSPHGS 115

Db 322 VDVMSFGIVLYLVCGKY-PRDQNMALAKIKKGVPEPY-----LSDCKGI 371

QY 116 QOSRKVDP-YHSTYMOQLS-----IRN-EGEPHLAGDRPSKHLKLVIRICLRALRLRP 169

Db 372 LSRMLVDPDKRAKLEEVKLNHPMIRNTEGPPSPAPERSP----- 412

QY 170 ISIDMESPLNLGSEALSPATAKDELITOMILKSAARS-ELGMY-----VSKROEFYL 222

Db 413 IILPDELIELIRNMGDFGPPREKIYRELTYKISEAVOSIAKTGFYSGPNSADKKSFEE 472

QY 223 RRARRRRKPAKPVLOSISEMKPVMEFTPM-----AYRDSGPPKNSRP--S 269

Db 473 FRIRAAHDIENILSLSMNTDIDAFHPLISITYIVSERRYTEKGSNNRIAKTIVSS 532

QY 270 LDPGNISPPROQVSVFQRRSPPPKNSVPPQPAFVARTASKYSAASQOVQRRNGNAK 326

Db 533 VP-----SSPVQ-----PTSNKRLTPMPEVY--AYKDESPVSRNLSLAR 573

RESULT 6

A54423

brevican precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jan-2000

C:Accession: A54423; S41914

R:Yamada, H.; Watanabe, K.; Shimonaka, M.; Yamaguchi, Y.

J. Biol. Chem. 269, 10119-10126, 1994

A:Title: Molecular cloning of brevican, a novel brain proteoglycan of the aggrecan/vers

A:Reference number: A54423; MUID:94193597

A:Accession: A54423

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-912 <YAM>

A:Cross-references: GB:X75887; NID:g452820; PIDN:CAA53481.1; PID:g452821

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology.

F:50-139/Domain: Immunoglobulin homology <IMM>

F:174-251/Domain: Link protein repeat homology <LNK1>

F:272-353/Domain: Link protein repeat homology <LNK2>

F:651-682/Domain: EGF homology <EGF>

F:689-809/Domain: C-type lectin homology <LCH>

F:816-872/Domain: complement factor H repeat homology <FHD>

Query Match 6.7%; Score 116; DB 2; Length 912;

Best Local Similarity 24.2%; Pred. No. 0.66;

Matches 54; Conservative 26; Mismatches 101; Indels 42; Gaps 8;

QY 140 PEPHLAGDRSKHLKLVIRICLRALRLP-----ISIDME-----SPILNLS 183

Db 368 PASHLASDALE--AIVTVEETLEELKLPQEAWESESGAIVSIPILEDGGSSSTPBDP 424

QY 184 GEALSPATATAKDEITOMILKSAARSELGMYVSKROEFLRRARRRR-----FA 232

Db 425 AEA--PRTLEFE-TQSIVPPILSSSEEGKVLQGEKTRGEEKEEPEEVEDEALMA 481

QY 233 WKPYLOSISEMKPVMEFTPMAYRDSGPPKNA-----STPSLPGKNSPPROQVSVQ 286

Db 482 WPSLESLDPEALPTEPVEPESITQASPPVRAALQPGVSPPEDEAPRRPVLGAPT 541

QY 287 RSSPPPKN--VSPPOPAFVARTASKYSAASQOVQRRNGNAK 326

Db 542 KTLPTREGNLASPPPTLVGAREIBETGPELSCAPRGSE 584

RESULT 7

P84643

hypothetical protein At2g25050 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: P84643

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shree, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

ems; D.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A64420; MUID:20083487

A:Accession: P84643

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-742 <STO>

A:Cross-references: GB:AE002093; NID:g4559347; PIDN:AAD23008.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g25050

A:Map position: 2

Query Match 6.6%; Score 114; DB 2; Length 742;

Best Local Similarity 21.7%; Pred. No. 0.72;

Matches 60; Conservative 48; Mismatches 127; Indels 42; Gaps 9;

QY 52 TITRNAAEVLEYGVTEFCATPNAOPFLERQPKVHRSKILPFRSKGHVRELSP 111

Db 352 TVFQITANIIQESLDGSPRSDSLSEALEKVEKTKM-----ISENVSP 404

QY 112 THGQSKRVFDYHSVTYMOQLSETRNDEPEPHLAGDRPSKHLKLVIRICLRALRLPIS 171

Db 405 DTSSPEKEK-----DTMSSHKSYAD--PNSILKRVDSRSLRVQVRNHSKITSPPMV 456

QY 172 IDLMESEPLNLGSEALSPATAKDELITOMILKSAARSELGMYVSKROEFLYRRA 225

Db 457 QSPVTSPLPDRSPPTGSPASISRFH-----SSPSSIGITSLIHDHSGCKDEESTSS 508

QY 226 RRRKRAKPKPVLOSISEMKPVMEF--HTPMAYRDSGPPKNSSTPSLPGKNISPPRO 281

Db 509 PASPSISFLPTLHLPTSSQPKKASRQCPSPPRVHSNGPSEANVTSSPLRP-LKPLRI 567
 Oy 282 VSPVORSSPPPKNNVSPPOPAFVARTASKYSASQOV 318
 Db 568 LSRP-----PPPPISLSLRSPSPSSSTNSI 596

RESULT 8

T38293

hypothetical serine-rich protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38293

R:Seillon, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, January 1996

A:Reference number: 221784

A:Accession: T38293

A:Status: preliminary; translated from GB/EMBL/DDAJ

A:Molecule type: DNA

A:Residues: 1-576 <SRP>

A:Cross-references: EMBL:268887; PIDN:CA93115.2; GSPDB:GN00066; SPDB:SPAC23E2.03c

A:Experimental source: strain 972h-; cosmid c23E2

C:Genetics:

A:Gene: SPDB:SPAC23E2.03c

A:Map position: 1

A:Introns: 553/2

Query Match 6.5%; Score 112.5; DB 2; Length 576;
 Best local similarity 23.8%; Pred. No. 0.68;
 Matches 75; Conservative 35; Mismatches 120; Indels 85; Gaps 14;

Oy 67 VTFVCAITNPAPFLERQPKVYHRSKILPRSKGV-HYRELSPRHSGQSKRVPEYH 125
 Db 18 ITRKICFEINA-----ATSPRSFOIQRLIKYAVIHGCFSLSHKHPTTASASDSD 69
 Oy 126 SYVMQOLESIRNEGPEPHLAGDRPSKHL-----KLVTRHCLR 163
 Db 70 EAV-----ELFTHSQPLPIPAQPSKSHADIFKFKPSKASLSLPCSKNSDVMNVCYMLK 125
 Oy 164 ALRLPRISI-----DLMESPLNLGSEALSPY-----ATAK-----DEITOMLIK 203
 Db 126 ATISKRYAVLSSGSAKAEILMWP-PNLAKRPLANSSEFSGSAKTQLEPEFSVY----P 180
 Oy 204 SAARSRLGMVYSKQEFYLRARRRRKFAKRPVLDISSEKMPVMEHTPM-----AYR 256
 Db 181 SASLVYQPSFNSNPAPITTSATHTSQF-----STSSSSSVNSVHTPVMPVNPYFDYN 233
 Oy 257 DSGSPKKNASTPLRCPKNISPPROVSVPORSSPP--PKNVSPPOPAFVARTASKYSAA 314
 Db 234 SSTATSSSSSSSVAP---FVPRROFSVSSASDPQPTPISMPPIP-----TPSQFSAF 284
 Oy 315 SOOVORNRGNASKLY 329
 Db 285 MYQNOGSYLSQSH 299

RESULT 9

T33658

hypothetical protein W07B3.2c - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33658

R:Wohlmann, P.; Bauer, C.; Rohlfing, T.; Giliam, B.

submitted to the EMBL Data Library, October 1999

A:Description: The sequence of C. elegans cosmid W07B3.

A:Reference number: 221384

A:Accession: T33658

A:Status: preliminary; translated from GB/EMBL/DDAJ

A:Molecule type: DNA

A:Residues: 1-539 <W0H>

A:Cross-references: EMBL:AF100304; PIDN:AMC68911.1; GSPDB:GN00021; CESP:W07B3.2c

A:Experimental source: strain Bristol N2; clone W07B3

C:Genetics:
 A:Gene: CESP:W07B3.2c
 A:Map position: 3
 A:Introns: 61/3; 101/3; 124/3; 257/3; 291/2; 364/3; 495/3

Query Match 6.4%; Score 111.5; DB 2; Length 539;
 Best local similarity 23.0%; Pred. No. 0.74;
 Matches 43; Conservative 38; Mismatches 75; Indels 31; Gaps 6;

Oy 154 KLVTRHCLRRLPRISID-LMESPLNLGSEALSTPATDEITOMLIKSAARSLCH 212
 Db 300 KRVYNQRTTEAFRRRREEDMLAMPICRIGLEAL-----DRAQITVRNKAAREALR 352
 Oy 213 YVSKROEFLYLRARRRRKFAKRPVLDISSEKMPVMEHTPMAYRDSGSPPKNASTPLRP 272
 Db 353 LRYQHMPDQKSYNQKRYTPKRRREMDMISAGQ---MAVGSGTITMGTGTPAKKD 408
 Oy 273 P-----KNISPPROVSVPOR---SSPPKVVSPPOPAFVARTASKYSASQO 317
 Db 409 DMDALSLERDYVKRQHAQQLLQGRANOSGTPRSYTPPOATITISNGQGH-----Q 464
 Oy 318 VORNRGN 324
 Db 465 LVGNQGN 471

RESULT 10

T01510

hypothetical protein T10M13.10 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 04-Mar-2000

C:Accession: T01510

R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hofman, J.; Hasegawa, A.; Gnoj, L.; G

Martensen, R.; McCombie, W.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.

A:Reference number: 214346

A:Accession: T01510

A:Status: translated from GB/EMBL/DDAJ

A:Molecule type: DNA

A:Residues: 1-202 <I0H>

A:Cross-references: EMBL:AF001308; NID:g2104523; PID:g2104533

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 45

A:Note: T10M13.10

C:Superfamily: Arabidopsis thaliana hypothetical protein T10M13.10

Query Match 6.4%; Score 111; DB 2; Length 202;
 Best local similarity 21.9%; Pred. No. 0.23;
 Matches 54; Conservative 24; Mismatches 89; Indels 80; Gaps 7;

Oy 16 RYVYPDGAVENLSKSCSYVDLLGNPDYVCGSTP-----YTTNRMA---AEVL 63
 Db 19 KVLVSDGRVQLEDETTVAEIMLENPOHVVEEDPSSISFNNDATVYRKALPLADKTL 78
 Oy 64 EYGVYVCAITNPAPFLERQPKVYHRSKILPRFSKHGVHVEILRSPHSGQSKRVPD 123
 Db 79 EBGKITYV-----LPA-KRSGRAAKSSAVLSEEMKML- 113
 Oy 124 YVSTVMOLESINNEGPPEPHLAGDRPSKHLKLVFIHCLRALRLRIDLMESPLNLS 183
 Db 114 -----FSATPAKRSSTFYBSILPWFT 135
 Oy 184 GSAISPAATAKDEITOMLIKSAARSELGMVYSKROEFLRARRRRKFAKRPVLDISEM 243
 Db 136 TRSYKNNNPMTD--TVVATSVGRLEAMEBEDRPEFLSROLSGR---GMPSLDPIREK 190
 Oy 244 KPVMEFH 250
 Db 191 KAKKKIH 197

Query Match 6.2%; Score 107; DB 2; Length 379;

Best Local Similarity 48.9%; Pred. No. 1;

Matches 22; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

RESULT 15

F96797

Probable G1-like trihelix DNA-binding protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F96797

R:Prologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Taiton,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F96797

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-603 <STO>

A:Cross-references: GB:AE005173; NID:g1054438; PIDN:AG27801.1; GSPDB:GN00141

C:Gene: F7012.5

C:Gene: F7012.5

A:Map position: 1

Query Match 6.2%; Score 107; DB 2; Length 603;

Best Local Similarity 17.0%; Pred. No. 1.9;

Matches 50; Conservative 54; Mismatches 88; Indels 102; Gaps 11;

OY 97 RFSKHGVHRELSPTHGSOOSRKVPDYHSVTMO-----OLESTRNG-- 139

DB 97 KMAEHGY-----IRNAKKCKEKFEVYKVKHRTKEGRGKSEGTYRFQLEALESOST 152

OY 140 -----PEPHLAGDRPSKHLKVF-----IRHCLRALPRISIDLMSPL 179

DB 153 SLHHHQOQTPRLPQOONNNNNNNNNNSTESTPTTPVTTPMLPSSSIPTTQOINVPSE 212

OY 180 PNLGSEALSPYATAKDETOMILKSARSELG-----MYVSK 216

DB 213 PNISGDFLSDNSTS-----SSSSTSTSDMEQCGATATRRKKRKMYFEERLKKQYVDK 268

OY 217 ROFE-----YLRARRR-----RFAWKPYLQISPMKVMVEFTPYAYDSCGPPKNA 267

DB 269 QEELQKKEFLAEVKEHEHRLVRESNR-----VQELAIRNREHILADERSMSAKDAV 323

OY 268 PSL-----PGKNISSPPROYS-----VPQSSPPPKNVSPPOP 301

DB 324 MAFLOKLSKOPNPOPOPOQOVPBSMOQNNNOQOPPOSSPPPOPAALPOP 377

Search completed: September 16, 2002, 00:44:16

Job time: 595 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2002, 00:44:21 ; Search time 24.18 Seconds
(without alignments)
530.032 Million cell updates/sec

Title: US-09-828-313-35

Perfect score: 1734

Sequence: 1 MGILPESCVTYQGYRVYYP.....SAASQOVRRGNKSLIYMA 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SWISSPROT_40.8

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	6.7	891	1	KIN1_SCHPO
2	116	6.7	912	1	PGCB_BOVIN
3	112.5	6.5	569	1	STRE7_SCHPO
4	105	6.1	2845	1	APC_MOUSE
5	103.5	6.0	830	1	DYNI_CAEEL
6	103	5.9	1822	1	ZAP3_HUMAN
7	103	5.9	2843	1	APC_HUMAN
8	103	5.9	3703	1	ABP1_HUMAN
9	102.5	5.9	864	1	WS14_MOUSE
10	102	5.9	620	1	EXTN_TOBAC
11	101.5	5.9	1321	1	IRS2_MOUSE
12	101	5.8	1540	1	RM1_DROME
13	100	5.8	1943	1	PC15_MOUSE
14	100	5.8	2842	1	APC_RAT
15	100	5.8	3726	1	ABP1_MOUSE
16	99.5	5.7	760	1	FXM1_MOUSE
17	99	5.7	555	1	GPI_CHIRE
18	99	5.7	759	1	FXM1_RAT
19	99	5.7	1021	1	YIB8_CAEEL
20	99	5.7	3530	1	MT15_HUMAN
21	98.5	5.7	806	1	MK07_MOUSE
22	98	5.7	457	1	WAS2_HUMAN
23	98	5.7	678	1	T2D5_MOUSE
24	97.5	5.6	1745	1	ZOI_MOUSE
25	97.5	5.6	306	1	EXTN_DAUCA
26	97	5.6	426	1	EXLP_TOBAC
27	97	5.6	630	1	T2D5_HUMAN
28	97	5.6	677	1	Y182_HUMAN
29	97	5.6	1157	1	ZAP3_MOUSE
30	96.5	5.5	1386	1	BLM_DROME
31	96	5.5	1487	1	RM1_HUMAN
32	96	5.5	1805	1	homo sapien
33	96	5.5	1805	1	homo sapien

34	95.5	5.5	564	1	ATR_HUMAN	09h622 homo sapien
35	94.5	5.4	522	1	POL2_RUBV	P08564 rubella vlr
36	94.5	5.4	701	1	CG1_HUMAN	O13495 homo sapien
37	94.5	5.4	732	1	POK_DROME	O01842 drosophila
38	94.5	5.4	787	1	SMO_HUMAN	O09835 homo sapien
39	94.5	5.4	815	1	MK07_HUMAN	O13164 homo sapien
40	94	5.4	597	1	V70K_OYAV	P20130 ononis ye11
41	94	5.4	1362	1	BRD4_HUMAN	O60885 homo sapien
42	94	5.4	2715	1	TRX2_HUMAN	O3um66 homo sapien
43	93.5	5.4	475	1	CAP1_HUMAN	O01518 homo sapien
44	93	5.4	395	1	KRP1_HUMAN	O06427 newcastle d
45	93	5.4	504	1	YHV9_YEAST	P38854 saccharomyc

ALIGNMENTS

RESULT 1
ID KIN1_SCHPO STANDARD: PRT: 891 AA.
AC P22987: 074392,
DI 01-AUG-1991 (Rel. 19, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase kin1 (EC 2.7.1.-).
GN KIN1 OR SPB4F6.06.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91045979; PubMed=2236039;
RA Levin D.E., Bishop J.M.;
RT "A putative protein kinase gene (kin1+) is important for growth
polarity in Schizosaccharomycetes pombe.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8272-8276(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gwilliam R., Rajadream M.A., Barrett B.G., Skelton J., Churcher C.M.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: PROBABLE SERINE/THREONINE PROTEIN KINASE. IMPORTANT FOR
CC GROWTH POLARITY IN S.POMBE.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STRONG, TO YEAST KIN1 AND KIN2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M64999; AAA3577.1; -
CC EMBL: AL031534; CAA20726.1; -
CC PIR: A38903; A38903.
CC HSSP: P00518; 1PHK.
CC InterPro: IPR000719; Euk-kinase.
CC InterPro: IPR001772; KAI.
CC InterPro: IPR002290; Ser_thr-kinase.
CC InterPro: IPR001245; Tyr-kinase.
CC Pfam: PF02149; KAI.1.
CC Pfam: PF00069; kinase.1.
CC PRINTS: PR00109; TYRKINASE.
CC SMART: SM00220; S_TKc.1.
CC PROSITE: PS00107; PROTEIN_KINASE_APP.1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST.1.
CC PROSITE: PS00111; PROTEIN_KINASE_DM.1.
CC Transferase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 125 395 PROTEIN KINASE.

[illegible][illegible]

OY 184 GEALSPATKDEITOMILKSANSELGMVSKROEFLRARRRK-----FA 232
 DB 425 AEA--PRTLEEF--QOSTIPPLGSSSEEGKVLGEQEKYNGEKEKEEVEEDALMA 481
 OY 233 WKPVLOSISMPPVMEFTPMAYRDSGSPKNA-----STPSLPGKKNISPROVSVQ 286
 DB 482 WPSLSLDEAPALPPEEVEEESTLQNSPPYRALQPGVSPPEYDEPADRPVPLGPT 541
 OY 287 RSSPPKKN--VSPPPQAPVARTASKRYSAASOOVQRNKNK 326
 DB 542 KTLPTREGNLASPPSTLVGAREIEETGPELSPGAPRGSE 584

RESULT 3

STEF_SCHPO STANDARD: PRT: 569 AA.
 ID Q10136:
 AC 01-FEB-1996 (rel. 33, Created)
 DT 01-FEB-1996 (rel. 33, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Ste7 protein.
 GN STE7 OR SPAC23E2.03C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxId=4896;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=20296734; PubMed=10835379;
 RA Matsuyama A., Yabana N., Matanabe Y., Yamamoto M.;
 RT "Schizosaccharomyces pombe Ste7p is required for both promotion and
 RT withholding of the entry to meiosis.";
 RN Genetics 155:539-549(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RA Skelton J., Church C.M., Barrell B.G., Rajadream M.A., Walsh S.V.;
 RT Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: HAS A ROLE IN PROMOTING MEIOSIS WHEREBY IT
 CC IS INVOLVED IN ESTABLISHING THE MATING PHEROMONE SIGNALING
 CC PATHWAY. IT ALSO HAS A ROLE IN SUPPRESSING MEIOSIS UNTIL THE
 CC CONJUGATION PROCESS IS COMPLETE.
 CC -!- INDUCTION: BY NITROGEN STARVATION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB036789; BAA90541.1; -
 DR EMBL: Z68887; CA93115.3; -
 DR Conjugation; Meiosis; Pheromone.
 FT DOMAIN 200 261
 FT DOMAIN 464 536 SER-RICH.
 FT SEQUENCE 569 AA; 61101 MW; FDD238939A7A6AC CRC64;

Query Match 6.58; Score 112.5; DB 1; Length 569;
 Best Local Similarity 23.8%; Pred. NO. 0.37;
 Matches 75; Conservative 35; Mismatches 120; Indels 85; Gaps 14;

OY 67 VYFYVQAPNAPLEPRQPVVHRSKILPRFSKGV-HVNEALSPTHSQOSRKAVPYH 125
 DB 18 ITKICIFETNA-----ATSPRSFOQLRKGYAVYHNGFLSSPSKHNPITASASD 69
 OY 126 SYVMQOLSEINNEGPBHLGDRPKHL-----KLVFRRHCLR 163
 DB 70 EAT-----ELFTHSQPLPIPAQPSKSHALDFKFKPSPSKASLSPCKSKSDMSWNVICYMLK 125

OY 164 ALALPRISL-----DLMESPDLNLSGALSTP-----ATAK-----DEITOMILK 203
 DB 126 ATISKRYAFLAGSSGASAKALIMP--PILAGKPLSANSFGSSAKTKTQLEDFDSVT----P 180
 OY 204 SAARSELGMVSKROEFLRARRRRKFAMKPVLOSISEMKPMVEFHTPM-----AYR 256
 DB 181 SASLYKQPSFNINPAPITTSATHTSQF-----STSSSSSVNSVHTVWVMPNRYQYN 233
 OY 257 DSGSPKKNASPSLPGKKNISPROVSVQBSPP--PKVSPPPQAPVARTASKRYSA 314
 DB 234 SSMTPASSSSSSVAP-----FVPRQPSVSSASDPQPTISKSPIPP-----TPSQSFAP 284
 OY 315 SOQVQRNKNKSLY 329
 DB 285 MYQNDSTYLSPOSHY 299

RESULT 4

APC_MOUSE STANDARD: PRT: 2845 AA.
 ID APC_MOUSE
 AC 061315: 062044;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein) (MAPC).
 GN APC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOPORMS 1 AND 2), AND VARIANTS.
 RX STRAIN=C57BL/6J AND CAST/ET; TISSUE=Brain;
 RX MEDLINE=92263101; PubMed=1350108;
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.;
 RT "Multiple intestinal neoplasia caused by a mutation in the murine
 RT homolog of the APC gene.";
 RL Science 256:668-670(1992).
 RN [2]
 RP ERRATUM.
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.;
 RL Science 256:1114-1114(1992).
 RN [3]
 RP SEQUENCE OF 1-45 FROM N.A.
 RX STRAIN=BALB/C; TISSUE=Liver;
 RA Dicker F., Lambert S., Reilmair A., Ballhausen W.G.;
 RT "The murine APC gene: alternative splicing of 5' untranslated
 RT region segments.";
 RL Submitted (OCT-1993) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=94061824; PubMed=8242607;
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
 RT "APC gene messenger RNA: novel isoforms that lack exon 7.";
 RL Cancer Res. 53:5589-5591(1993).
 CC -!- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
 CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY
 CC SIMILARITY).
 CC -!- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
 CC SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: 4 ISOPORMS: 1 (SHOWN HERE), 2, 3 AND 4; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
 CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
 CC -!- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 7 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL:	M88127;	AAB59632.1;	-
DR	EMBL:	U02937;	AAA03443.1;	-
DR	HSP:	O02248;	3BCT.	
DR	MGI:	M88039;	APC.	
DR	InterPro:	IPR000225;	Armadillo.	
DR	Pfam:	PF00514;	Armadillo_seg. 6.	
DR	SMART:	SM00185;	ARM. 5.	
DR	PROSITE:	PS50176;	ARM_REPEAT. 1.	
KW	Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;			
KM	Coiled coil.			
FT	DOMAIN	1	245	COILED COIL (POTENTIAL).
FT	DOMAIN	125	61	COILED COIL (POTENTIAL).
FT	REPEAT	1	728	LEU-RICH.
FT	REPEAT	451	493	ARM 1.
FT	REPEAT	503	545	ARM 2.
FT	REPEAT	546	589	ARM 3.
FT	REPEAT	590	636	ARM 4.
FT	REPEAT	637	681	ARM 5.
FT	REPEAT	682	723	ARM 6.
FT	REPEAT	724	765	ARM 7.
FT	DOMAIN	739	2834	SER-RICH.
FT	DOMAIN	1130	1156	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	1556	1575	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	1864	1891	HISLY CHARGED.
FT	DOMAIN	243	276	MISSING (IN ISOFORM 2 AND ISOFORM 4)
FT	VARSPLIC	310	410	MISSING (IN ISOFORM 3 AND ISOFORM 4)
FT	VARIANT	120	120	T -> A (IN STRAIN CAST/EI).
FT	VARIANT	493	493	Y -> I (IN STRAIN CAST/EI).
FT	VARIANT	797	797	V -> F (IN STRAIN CAST/EI).
FT	VARIANT	1330	1330	A -> T (IN STRAIN CAST/EI).
FT	VARIANT	1638	1638	A -> S (IN STRAIN CAST/EI).
FT	VARIANT	2294	2294	G -> A (IN STRAIN CAST/EI).
FT	VARIANT	2436	2436	H -> A (IN STRAIN CAST/EI).
FT	VARIANT	2532	2532	T -> Q (IN STRAIN CAST/EI).
FT	VARIANT	2813	2813	T -> S (IN STRAIN CAST/EI).
SO	SEQUENCE	2845 AA;	311086 MW;	14SCA3CE570A459 CRC64;

[illegible]


```
Db      664  MRLTKITIKLVKRAVMHL-----IVNCTGEFPM-----KDELLAHLYCGSDTALMEESQ 713
QY      212  MYVSROEFLRRARRRRKFAWKPYLQSTSE--MKPYMEFHPMAVYRSGSPPKNASTPS 269
Db      714  IEACKREEML-----RMTHACKELALPITSYVNMSTIGQPPPLPMASDYPRHPSGSPVPV 767
QY      270  LFGPKNTISPP--RQVSVPOR---SSPPPKNVSPPP 299
Db      768  RPAP---APPGGRQAPMPRGAPPPGAPPPGMRPP 799

RESULT 6
ZAP3_HUMAN
ID      ZAP3_HUMAN      STANDARD:      PRT:      1822 AA.
AC      P49750; P49752; 09P1V7;
DT      01-OCT-1996 (Rel. 34, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      Nucleat protein ZAP3 (ZAP113).
GN      ZAP3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RA      SEQUENCE FROM N.A.
RA      Rowen L., Madan A., Qin S., Abbsasi N., Baradaran L., Birditt B.,
RA      Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA      Kaur A., Maan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RA      "Sequencing of human chromosome 14q24.3 region."
RA      Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN      SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=95319502; PubMed=7596406;
RA      Sherrington R., Rogaeve E.I., Liang Y., Rogaeve E.A., Lavesque G.,
RA      Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA      Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA      Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA      Sansen P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA      Petrick-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA      Rommens J.M., St George-Hyslop P.H.;
RA      "Cloning of a gene bearing missense mutations in early-onset familial
RA      Alzheimer's disease."
RT      Nature 375:754-760(1995).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC      -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC      frameshift in position 1661.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch)
-----
DR      EMBL: AC007956; AAF61275.1; -
DR      EMBL: L40403; AAC42008.1; ALT_FRAME.
DR      EMBL: L40400; AAC42006.1; -.
KW      Nuclear protein.
FT      DOMAIN 15 205 PRO-RICH.
FT      DOMAIN 382 430 GLN-RICH.
FT      DOMAIN 807 1209 ARG-RICH.
FT      DOMAIN 1488 1577 ARG-RICH.
FT      CONFLICT 621 621 P -> S (IN REF. 2).
FT      CONFLICT 1404 1404 T -> I (IN REF. 2).
FT      CONFLICT 1821 1821 K -> E (IN REF. 2).
SQ      SEQUENCE 1822 AA; 204947 MW; 8E6C83FE540C7D2 CRC64;
```

Query Match 5.9%; Score 103; DB 1; Length 1822;

```
Best Local Similarity 25.1%; Pred. No. 8.1;
Matches 47; Conservative 30; Mismatches 76; Indels 34; Gaps 9;
QY      159  RHCIKRLRLP--RISIDMESPLPNSLGEALSPFATANDETITOMILKSAARSLGMYSK 216
Db      1217  RHSPMEHNPSSHHSHSEMGSDASIDSDGIGVWVLSQRQHEITLKA--QELKMLRQ 1274
QY      217  ROEFLRRARRRRKFAWKPY-----LQSTSEMKPYMEFHPMAV---DSGSP 262
Db      1275  KEQI-----QMKKDGSEPGMAADLPQESRLQNTSS--RPGM--YPPGSTRPPPMKAP 1327
QY      263  KNASTSLPGPKNISPPRQVSVPORSSPPPKNVSPPPQAPAFVARTASKYSAAQVQGNR 322
Db      1328  GSTVPSAPPARSSVPTVRPPVPIPPPPPLP-PPPPVTKPQTSN-----VEQER 1378
QY      323  GNKASLY 329
Db      1379  WDEDSFY 1385
```

```
RESULT 7
APC_HUMAN
ID      APC_HUMAN      STANDARD:      PRT:      2843 AA.
AC      P25054; Q15162; Q15163;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Adenomatous polyposis coli protein (APC protein).
GN      APC OR DP2.5.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=91335210; PubMed=1651562;
RA      Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
RA      Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
RA      Finlind R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,
RA      Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.;
RA      "Identification of FAP locus genes from chromosome 5q21."
RA      Science 253:661-665(1991).
[2]
RN      SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RC      TISSUE=Fetal brain;
RX      MEDLINE=91330307; PubMed=1678319;
RA      Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,
RA      Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA      Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
RA      Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
RA      Adzharihan H., Cohen D., Leppert M., White R.;
RA      "Identification of deletion mutations and three new genes at the
RA      familial polyposis locus."
RA      Cell 66:601-613(1991).
[3]
RN      ASSOCIATION WITH CATENINS.
RP      MEDLINE=94082295; PubMed=8259519;
RA      Su L.-K., Vogelstein B., Kinzler K.W.;
RT      "Association of the APC tumor suppressor protein with catenins."
RA      Science 262:1734-1737(1993).
[4]
RN      REVIEW ON VARIANTS.
RP      MEDLINE=94154728; PubMed=8111410;
RA      Nagase H., Nakamura Y.;
RT      "Mutations of the APC (adenomatous polyposis coli) gene."
RA      Hum. Mutat. 2:425-434(1993).
[5]
RN      VARIANTS FAP.
RX      MEDLINE=91335211; PubMed=1651563;
RA      Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
RA      Koyama K., Utsunomiya Y., Baba S., Hedge P., Markham A., Krush A.J.,
RA      Petersen G., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA      Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
```

RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.";
 RL Science 253:665-669(1991).
 RN [6]
 RP VARIANTS FAP.
 RX MEDLINE-93265030; PubMed-1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.;
 RT "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.";
 RL Hum. Mol. Genet. 1:229-233(1992).
 RN [7]
 RP VARIANTS FAP.
 RX MEDLINE-93244793; PubMed-1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.;
 RT "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.";
 RL Hum. Mol. Genet. 1:559-563(1992).
 RN [8]
 RP VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 AND S-2502.
 RX MEDLINE-93250848; PubMed-1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
 RA Baba S., Nakamura Y.;
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RT patients: 61 new patients and a summary of 150 unrelated patients.";
 RL Hum. Mutat. 1:467-473(1992).
 RN [9]
 RP VARIANT FAP TRP-99.
 RC TISSUE-Peripheral blood lymphocytes;
 RX MEDLINE-95134544; PubMed-7833149;
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.;
 RT "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.";
 RL Eur. J. Cancer 30A:1709-1713(1994).
 RN [10]
 RP VARIANT FAP GLY-722.
 RX MEDLINE-9513530; PubMed-7833931;
 RA Stella A., Montero M., Reata N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilla S., Prete F., Mareni C., Guanti G.;
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.";
 RL Hum. Mol. Genet. 3:1687-1688(1994).
 RN [11]
 RP ERRATUM.
 RA Stella A., Montero M., Reata N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilla S., Prete F., Mareni C., Guanti G.;
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [12]
 RP VARIANT FAP IIE-171.
 RX MEDLINE-97144176; PubMed-8990002;
 RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.;
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
 RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGGE, PTT, and southern analysis.";
 RL Hum. Mutat. 9:7-16(1997).
 RN [13]
 RP VARIANT LYS-1307.
 RX MEDLINE-98400248; PubMed-9731522;
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhansen S.L., Satagopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Oberis D., Andrulis I., Daly M., Plinsky L., Schrag D.,
 RA Goldinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
 RT "The APC I1307K allele and breast cancer risk.";
 RL Nat. Genet. 20:13-14(1998).
 RN [14]
 RP VARIANTS LYS-1307 AND GLN-1317.

RC TISSUE-Peripheral blood;
 RX MEDLINE-98393712; PubMed-9724771;
 RA Freyling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
 RT "The APC variants I1307K and E1317G are associated with colorectal
 RT tumors, but not always with a family history.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [15]
 RP VARIANT LYS-1307.
 RX MEDLINE-98400259; PubMed-9731533;
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
 RT "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews.";
 RL Nat. Genet. 20:62-65(1998).
 RN [16]
 RP VARIANT LYS-1307.
 RX MEDLINE-99138651; PubMed-9973276;
 RA Gryfe R., Di Nicola N., Lai G., Gallinger S., Redston M.;
 RT "Inherited colorectal polyposis and cancer risk of the APC I1307K
 RT polymorphism.";
 RL Am. J. Hum. Genet. 64:378-384(1999).
 RN [17]
 RP VARIANTS FAP CYS-1171 AND THR-2738, AND VARIANTS GLY-1057 AND
 RP VAL-1822.
 RX MEDLINE-99133859; PubMed-9950360;
 RA Wallis Y.L., Morton D.G., McKeown C.M., MacDonald F.;
 RT "Molecular analysis of the APC gene in 205 families: extended
 RT genotype-phenotype correlations in FAP and evidence for the role of
 RT APC amino acid changes in colorectal cancer predisposition.";
 RL J. Med. Genet. 36:114-20(1999).
 RN [18]
 RP VARIANT FAP PRO-1184.
 RX MEDLINE-99401091; PubMed-10470088;
 RA Lamum H., Ilyas M., Rowan A., Clark S., Johnson V., Bell J.A.,
 RA Freyling I.M., Efstathiou J., Pack K., Payne S., Roylance R.,
 RA Gorman P., Sheer D., Neale K., Phillips R., Talbot I.C., Bodmer W.F.,
 RA Tomlinson I.P.M.;
 RT "The type of somatic mutation at APC in familial adenomatous polyposis
 RT is determined by the site of the germline mutation: a new facet to
 RT Knudson's 'two-hit' hypothesis.";
 RL Nat. Med. 5:1071-1075(1999).
 CC -1- FUNCTION: "TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
 CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ.
 CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATEININS.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B.
 CC -1- DISEASE: DEFECTS IN APC ARE A CAUSE OF FAMILIAL ADENOMATOUS
 CC POLYPOSIS (FAP) AND GARDNER'S SYNDROME (GS), THAT CONTRIBUTE TO
 CC TUMOR DEVELOPMENT IN PATIENTS WITH NONINHERITED FORMS OF
 CC COLORECTAL CANCER. FAP IS CHARACTERIZED BY ADENOMATOUS POLYPS OF
 CC THE COLON AND RECTUM, BUT ALSO OF UPPER GASTROINTESTINAL TRACT
 CC (APBULANT, DUODENAL AND GASTRIC ADENOMAS)..THIS IS A VICIOUSLY
 CC PERMANENT DISEASE WITH ONE OR MORE POLYPS PROGRESSING THROUGH
 CC DYSPLASIA TO MALIGNANCY IN UNTREATED GENE CARRIERS WITH A MEDIAN
 CC AGE AT DIAGNOSIS OF 40 YEARS.
 CC -1- DISEASE: APC MUTATIONS HAVE LED TO SOME INTERESTING OBSERVATIONS.
 CC (1) THE GREAT MAJORITY OF THE MUTATIONS FOUND TO DATE WOULD RESULT
 CC IN TRUNCATION OF THE APC PRODUCT. (2) ALMOST ALL THE MUTATIONS
 CC HAVE OCCURRED WITHIN THE FIRST HALF OF THE CODING SEQUENCE, AND
 CC SOMATIC MUTATIONS IN COLORECTAL TUMORS ARE FURTHER CLUSTERED IN A
 CC PARTICULAR REGION, CALLED MCR (MUTATION CLUSTER REGION). (3) MOST
 CC IDENTIFIED POINT MUTATIONS IN THE APC GENE ARE TRANSITIONS FROM
 CC CYTOSINE TO OTHER NUCLEOTIDES. (4) THE LOCATION OF COLORECTAL POLYPS
 CC MUTATIONS TENDS TO CORRELATE WITH THE NUMBER OF COLORECTAL POLYPS
 CC IN FAP PATIENTS. INACTIVATION OF BOTH ALLELES OF THE APC GENE
 CC SEEMS TO BE REQUIRED AS AN EARLY EVENT TO DEVELOP MOST ADENOMAS
 CC AND CARCINOMAS IN THE COLON AND RECTUM AS WELL AS SOME OF THOSE IN
 CC THE STOMACH.


```

FT DOMAIN 3376 3389 POLY-GLN.
FT DOMAIN 3392 3395 POLY-GLN.
FT DOMAIN 3507 3527 POLY-GLY.
FT DOMAIN 3597 3600 POLY-PRO.
FT DOMAIN 3636 3639 POLY-SER.
FT VARSPLC 1 3914 MISSING (IN ISOFORM B).
FT VARIANT 3374 3374 A -> V.
FT VARIANT 3374 3374 A -> V.
FT VARIANT 3377 3384 /fTid=VAR.011694.
FT VARIANT 3377 3384 MISSING.
FT VARIANT 3527 3527 /fTid=VAR.011695.
FT VARIANT 3527 3527 G -> GGG.
FT CONFLICT 72 72 /fTid=VAR.011696.
FT CONFLICT 422 422 A -> S (IN REF. 3).
FT CONFLICT 579 579 A -> P (IN REF. 3).
FT CONFLICT 767 767 T -> A (IN REF. 3).
FT CONFLICT 777 777 I -> S (IN REF. 3).
FT CONFLICT 846 849 A -> V (IN REF. 3).
FT CONFLICT 927 927 HHRV -> RHLG (IN REF. 3).
FT CONFLICT 997 997 A -> S (IN REF. 3).
FT CONFLICT 1150 1190 EALVDGSEFTADPEAPLAKNDQEGASSQAEKELTDS
FT CONFLICT 1150 1190 -> GEMSHRGRRRLGCHLLETLSRGLTEGSDVDPAPH
FT CONFLICT 1150 1190 VPT (IN REF. 3).
SQ SEQUENCE 3703 AA: 404468 MM: 0f62AF37D4DCf856 CRC64:
VPT (IN REF. 3).

Query Match
Best Local Similarity 2B.0%; Score 103; DB 1; Length 3703;
Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;

OY 171 SIDLE-----SPLPNLSGELSPFR--TAKETIOTMKSASARETGMYSKRQ 218
Db 2368 SMDAMEILTPTSSCSCTPMDSQYSAAPASNNNTASSAFLOL---TAAEELATFNKSTE 2424
OY 219 EYFLRARRRRKKRFM-----KPLQSIISMKRVMEFHPPMAYRDSOSP 262
Db 2425 -----ACGEKRLAEASAPQNTQERKGQCRPELO--QGEDEQKXTNTPQO-----KLP 2472
OY 263 KNASTPSLPGRKNISIPROVSYVDQNSPPPKXNSPP 299
Db 2473 QIVSLPSLPQPPQAPPPQPCPLPQ--SSPSPOLSHLP 2508

RESULT 9
ID WS14_MOUSE STANDARD: PRT: 864 AA.
AC 099MW3: 099MW0: 099MZ1: 099WZ2: 099WY9;
AD 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Williams-beuren syndrome chromosome region 14 protein homolog (MIX
DE Interactor).
GN WBSK14 OR MIO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=21153101; PubMed=11230181.
RX CALTO S., Merla G., Urbinali F., Ballabio A., Raymond A.,
FT "WBSK14, a gene mapping to the Williams-beuren syndrome deleted
FT region, is a new member of the MIX transcription factor network.",
RL Hum. Mol. Genet. 10:617-627(2001).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR. BINDS TO THE CANONICAL AND
CC -1- NON-CANONICAL E BOX SEQUENCES 5'-CAGCG-3' (By similarity).
CC -1- SUBUNIT: BINDS DNA AS A HETERODIMER WITH TCFL4/MX (BY
CC -1- SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1/ZETA (SHOWN HERE), 2/THETA,
CC 3/IOTA, 4/KAPPA AND 5/ETA: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.

```

[illegible]

```

RC STRAIN=CV, XANTHI; TISSUE=Leaf;
RC MEDLINE=90128263; Pubmed=2612909,
RT Kellier B., Lamb C.J.:
RT "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.",
RT Genes Dev 3:1639-1646(1989).
CC -I- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -I- SUBCELLULAR LOCATION: Extracellular matrix.
CC -I- PWM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13885; CA332090.1; -
DR PIR: S06733; S06733.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 620 EXTENSIN.
FT REPEAT 70 73 H-A-P-P.
FT REPEAT 148 151 H-A-P-P.
FT DOMAIN 229 242 H-A-P-P.
FT REPEAT 229 235 2 x 7 AA TANDEM REPEATS OF T-H-R-A-P-P.
FT REPEAT 236 242 1.
FT REPEAT 236 242 2.
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN 499 600 3 x APPROXIMATE TANDEM REPEATS.
FT DOMAIN 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
SO SEQUENCE

Query Match 5.9%; Score 102; DB 1; Length 620;
Best Local Similarity 32.4%; Pred. No. 2.5;
Matches 23; Conservative 5; Mismatches 31; Indels 12; Gaps 1.

QY 245 PWFMEHTPMAYRDSGSPPKNASTPSLPSPKNISPP-----ROYSVPRSSSPP 292
DB 378 PPFSPSPPTPEYEOSPPPPAVSPPLAPLPPTSPSPPTSPSPPTVAQPPPLPPTSPSP 437
QY 293 KNVSPPOPAF 303
DB 438 PAYSPPPPPT 448

RESULT 11
IRS2_MOUSE
ID IRS2_MOUSE STANDARD; PRF: 1321 AA.
AC P81122;
DT 13-JUL-1998 (Rel. 36, Created)
DT 13-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Insulin receptor substrate-2 (IRS-2) (4PS).
GN IRS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=95405472; Pubmed=7675087.
RA Sun X.J., Wang L.-M., Zhang Y., Yensh L., Myers M.G., Jr.,
RA Glaeser E., Lane W.S., Pierce J.H., White M.F.:
RA "Role of IRS-2 in insulin and cytokine signaling.",
RA Nature 377:173-177(1995).
CC -I- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN.

```

CC	-1	TISSUE SPECIFICITY: SKELETAL MUSCLE, LUNG, BRAIN, LIVER, KIDNEY,
CC	HEART AND SPLEEN.	
CC	-1 SIMILARITY: CONTAINS 1 PH DOMAIN.	
CC	-1 SIMILARITY: CONTAINS 1 PTB DOMAIN.	
CC	HSSP; P35568; 11RS.	
DR	M6D; MG1:109334; ITS2.	
DR	InterPro; IPRO02404; Insulin_Recep_S-1.	
DR	InterPro; IPRO01849; PH.	
DR	Pfam; PF02174; IRS; 1.	
DR	Pfam; PF00169; PH; 1.	
DR	PRINTS; PR00628; INSULINRSI.	
DR	SMART; SM00233; PH; 1.	
DR	SMART; SM00310; PTB; 1.	
DR	PROSITE; PS50003; PH_DOMAIN; 1.	
KM	Phosphorylation.	
FT	DOMAIN 16 144 PH.	
FT	DOMAIN 187 300 PTB.	
FT	MOD_RES 536 536 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 649 649 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 671 671 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 911 911 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 970 970 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 1242 1242 PHOSPHORYLATION (BY INSR)	
FT	MOD_RES 1303 1303 PHOSPHORYLATION (BY INSR)	
FT	DOMAIN 19 28 POLY-ASN.	
FT	DOMAIN 444 449 POLY-SER.	
FT	DOMAIN 638 641 POLY-SER.	
FT	DOMAIN 936 939 POLY-SER.	
SO	SEQUENCE 1321 AA; 136526 MW; 50699CE9D614960C7 CNC64;	
Query Match	5.9%; Score 101.5; DB 1; Length 1321;	
Best Local Similarity	21.4%; Pred. No. 7;	
Matches 78; Conservative 43; Mismatches 131; Indels 113; Gaps 18;		
QY	27 LSKSCSYHDLIGNPDYVYGSTPTTITNRMAEEVLEYGVTVYCATTAPPAQFLEROPK 86	
DB	336 LRSR--HTLSAG-----CGGRPSKYT-LAPAGALQH-----SRNSMPVAHSPRA 438	
QY	87 VNHGSKILPRRSKHG-----VHRELRSPTHGSOOS-----RK 120	
DB	439 ATSPGS--LSSSGHSGSGYRLPPGSHPLRPHLHNGQGRSSGSAASGSPDPGPM 496	
QY	121 VEDYHNV--TMQLESTRNGCPR-----HLAGDPRSHIKLVTIFHC 161	
DB	497 IDEGSGPFDLAFSHRSNTPTPSIAETPRPADSGGELGTGMSMDR-----LSHC 548	
QY	162 LRALRLPRISIDLMESPLMUSEALSPATAKDETOMILKSAARELDQNTVSKROEY 221	
DB	549 GRPYR--RVSGGAGADDRLGLRRRTYSLTPAR---ORQVPOSSASLDETTLKRAFS 602	
QY	222 LRRAR-----RRRKFAKMPVLQISIEK-----PYMEHTPMAVR 256	
DB	603 GSSRLCPSPHPASSKVAAYNPVEDYGDIEIGHSKSSSNLGDADGCTMPV---TPGAL 658	
QY	257 DSGSPKMASTSLP-LGKNISPPROVSPQSPRP--PKNVSPPPQAFAPART-----ASK 310	
DB	659 RSGGPNKSCDDYDMNMTPTSVSAKPKLQDPLAALPRGGAIVAPRPSGVGRFPVYNGG 718	
QY	311 YSAAS 315	
DB	719 YKASS 723	

```

ID  RM1.DROME          STANDARD;          PRT;          1540 AA.
AC  09V7M4;
DT  16-OCT-2001 (rel. 40, Created)
DT  16-OCT-2001 (rel. 40, Last sequence update)
DT  16-OCT-2001 (rel. 40, Last annotation update)
DE  RM1 Protein homolog.
GN  CG8370.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
XX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BRKLEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
RA  Abril J.F., Adayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.V., Bencze P.V., Bernier B.P., Bhandari D., Bolintinas S.,
RA  Borovaya D., Botchan M.R., Bouck J., Brodeur G., Cantu E., Center A., Chandra I.,
RA  Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferriz C., Ferrelia S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Meltel B., McIntosh T.C., McLeod M.P., McPherson A.,
RA  Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclob J.M.,
RA  Palazolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA  Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RN  Science 287:2185-2195(2000).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE003808; AAF58081; .
DR  FLYBase; FBgn0034060; CG8370.
DR  InterPro; IPR001121; G_adaptC.
DR  Pfam; PF02883; Alpha_adaptC2; 1.
SQ  SSOURCE 1540 AA; 170274 MW; 5ABCG7531D425846 CRC64;

```

```

QY 121 VEDYHSVTMOQ--LESTRNEGPEPHLADRPDKHLKLVFTRICRALRLPRISIDLMESP 178
DB 1040 VEDAKAIWVQONAYDAARNKGP----- 1061
QY 179 LPLSLGEALSPPTAVAKDEITOMILKSARSELGMVSKROEFLRRARRR----- 228
DB 1062 -----LQTFPLRNIVLQAEFAAKAESVOOQKVKNCQJKEFLRRRTVYNSTNSKS 1113
QY 229 -KKFAAKP-----VQSSSEBKPYMEFHPMAVROSGSP-----KNASTP 268
DB 1114 KSKSWSPMSMDNALSKHLQAKPRTVSTVTPPPAASAPAAAPPLPEAKPVKKSSTP 1173
QY 269 SLPG-----PKNISP-----PRO-VSP-----QSSSP-PKNNSP 298
DB 1174 SPQGVATISVQVAPQKVKPPTAVVLCTTKPKQEVSTPVADQHEKSLAKSSPQDENISPK 1233
QY 299 POPAFVARTASVKYSMAAGVOYRRNGNAK 326
DB 1234 PKNPPQRYLAKQNSAKKNGKTPGRER 1261

RESULT 13
PC15_MOUSE          STANDARD;          PRT;          1943 AA.
AC  099PUL;
DT  01-MAR-2002 (rel. 41, Created)
DT  01-MAR-2002 (rel. 41, Last sequence update)
DT  01-MAR-2002 (rel. 41, Last annotation update)
DE  Protocadherin 15 precursor.
GN  PCDH15.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  MEDLINE=20578757; PubMed=11138007;
RA  Alagramam K.N., Murcia C.L., Kwon H.Y., Pawlowski K.S., Wright C.G.,
RA  Woychik R.P.;
RT  "The mouse Ames waltzer hearing-loss mutant is caused by mutation of
RT  Pcdh15, a novel protocadherin gene.";
RN  Net. Genet. 27:95-102(2001).
RN  [2]
RN  TISSUE SPECIFICITY.
RX  PubMed=11429292;
RA  Murcia C.L., Woychik R.P.;
RT  "Expression of Pcdh15 in the inner ear, nervous system and various
RT  epithelia of the developing embryo.";
RN  Mech. Dev. 105:163-166(2001).
CC  -1- FUNCTION: Calcium-dependent cell-adhesion protein. Essential for
CC  maintenance of normal retinal and cochlear function. Required for
CC  inner ear neuroepithelial cell elaboration.
CC  -1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
CC  -1- TISSUE SPECIFICITY: Expressed in brain and sensory epithelium of
CC  the developing inner ear. Also found in the spleen, developing
CC  eye, dorsal root ganglion, dorsal aspect of neural tube, floor
CC  plate and ependymal cells adjacent to the neural canal.
CC  -1- DEVELOPMENTAL STAGE: Highest level of expression is detected at
CC  embryonic day 16.
CC  -1- DISEASE: Defects in PCDH15 are the cause of the Ames waltzer (av)
CC  phenotype. It is characterized by deafness and a balance disorder,
CC  associated with the degeneration of inner ear neuroepithelia.
CC  -1- SIMILARITY: CONTAINS 11 CADHERIN DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----

```

Query Match 5 8%; Score 101; DB 1; Length 1540;
 Best Local Similarity 18.7%; Pred. No. 9.2; 71; Indels 108; Gaps 10;
 Matches 50; Conservative 39; Mismatches

DR EMBL: AF281899; AAG53891.1; -
 DR HSSP: P09803; 1SDH.
 DR MGD: MGI:1891428; Pcdh15.
 DR InterPro: IPR002126; Cadherin.
 DR Pfam: PF00028; cadherin. 10.
 DR PRINTS: PR00205; CADHERIN.
 DR SMART: SM00112; CA. 11.
 DR PROSITE: PS00232; CADHERIN_1; 4.
 DR PROSITE: PS50268; CADHERIN_2; 11.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal: Multigene family; Vision.
 FT SIGNAL 1 26
 FT CHAIN 27 1943
 FT DOMAIN 27 1381
 FT TRANSMEM 1382 1402
 FT DOMAIN 1403 1943
 FT DOMAIN 145 152
 FT DOMAIN 153 270
 FT DOMAIN 283 400
 FT DOMAIN 401 514
 FT DOMAIN 515 621
 FT DOMAIN 622 722
 FT DOMAIN 724 824
 FT DOMAIN 825 931
 FT DOMAIN 932 1040
 FT DOMAIN 1042 1149
 FT DOMAIN 1150 1264
 FT DOMAIN 1437 1778
 FT DOMAIN 1772 1778
 FT DOMAIN 1804 1812
 FT CARBOHD 102 102
 FT CARBOHD 206 206
 FT CARBOHD 424 424
 FT CARBOHD 564 564
 FT CARBOHD 667 667
 FT CARBOHD 729 729
 FT CARBOHD 773 773
 FT CARBOHD 826 826
 FT CARBOHD 856 1069
 FT CARBOHD 1069 1089
 FT CARBOHD 1180 1180
 SQ SEQUENCE 1943 AA: 214816 MW: E3D7DB9F5F738652 CRC64;
 Query Match 5.8%; Score 100; DB 1; Length 1943;
 Best Local Similarity 23.7%; Pred. No. 15;
 Matches 63; Conservative 32; Mismatches 93; Indels 78; Gaps 12;
 QY 51 YTTTBMAREVL-----FYGVTFVCAFPNAPFLERQPKYVHRGSKILLPRSKHGCHV 105
 DB 1607 KQITWQAEESARCPHSGSSNVLLATEDAH---ESKKEGHRDILLVQO-----T 1656
 QY 106 RLNSLPTNGSQGRKVFYDHYSTMQOLESTRNEGPPHLAGDRSKHLKLVTRICLRAL 165
 DB 1657 EQLKLSLSSGSSFSNHSFTLPTJISRAVELGSSPNVYSPPADCTLEL-----SPPL 1710
 QY 166 RLPRISIDLMESPLPMLSGEALSPATATAKDEITOMILKSAARSELGMYVKROEFLRRA 225
 DB 1711 R-PR-----LNSLSKRETPCASCDE-----SPPL 1732
 QY 226 RRRKFAKPKVLOSISEMKPVMEFPTPMAYRDSGSPPKNASTPLP-GPKNISPPR--- 280
 DB 1733 PKRNFELAPHPSTIRAPLP---HPPLP-----RPLAFTTFPLPLSPNPDPQLVTF 1783
 QY 281 --QVSVPGRSS---PPKKNVSPROP 301
 DB 1784 SLPTSTPTSSLPLPPPLSLPPPPRP 1809

ID APC-RAT STANDARD: PRT: 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISCHER 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RL "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
 RL Mamm. Genome 6:746-748(1995).
 RN [2]
 RP MUTAGENESIS.
 RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
 RL -I- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
 ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY
 SIMILARITY).
 CC -I- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
 SIMILARITY).
 CC -I- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -I- SIMILARITY: CONSTRAINTS 7 ARM REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D38629; BAA07609.1; -
 DR HSSP: Q02248; 3BCT.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 6.
 DR SMART: SM00185; ARM; 5.
 DR PROSITE: PS50176; ARM_REPEAT; 1.
 DR Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
 FT DOMAIN 1 728
 FT DOMAIN 1 62
 FT DOMAIN 125 260
 FT REPEAT 451 493
 FT REPEAT 503 545
 FT REPEAT 546 589
 FT REPEAT 590 636
 FT REPEAT 637 681
 FT REPEAT 682 723
 FT REPEAT 724 765
 FT REPEAT 739 783
 FT DOMAIN 1130 1155
 FT DOMAIN 1556 1575
 FT DOMAIN 1864 1891
 FT MYRGEN 523 523
 SQ SEQUENCE 2842 AA: 310530 MW: 3CB2E8A34E8F47 CRC64;
 Query Match 5.8%; Score 100; DB 1; Length 2842;
 Best Local Similarity 21.9%; Pred. No. 23;
 Matches 59; Conservative 40; Mismatches 97; Indels 74; Gaps 12;


```
Dh 101 NYGSPVYVHNHNNYSSMSPSPMDPADRRQPEQARRLSVATDNMKLEFYKDKMGVYVMD 160
Qy 110 S---PTNGSQOSKRFEDHSVTYMQOLEISINEGPEPHLAGDPSKHLKLVIFIRHCLRALR 166
Dh 161 TNNWYARRGSSAGKRV-----SCAPSPMQPPAPPAELAAPLPSF----- 198
Qy 167 LPRISIDLMSPILNISEALSP-----TATAKDEITQMLISAAR----- 207
Dh 199 LPEQPLDSPAAPALPSVSLGLQPEPPEPTSTTKSK--ELSPGSAQKGSBSSQGTACACTIO 256
Qy 208 -----SELGMYVKRQEFYLRARRRRKFPAMKPVLOSISEMKPYMEFHTP--MAYRDSG 259
Dh 257 PGAGPQAGAPGASPSGSPADQSPHTLRK-----VSKKLAIPIPKVVFQGPAGMADQSAAG 311
Qy 260 SPPK---NASTPGLPEPKNISPPROVSVP-----QRSSPPPK- 293
Dh 312 QPSVSLSPPTPEPTSPICGLSPGCGYSLASGQLSPAAPPLASPSVFTSTLSKSRPTPKP 371
Qy 294 ---NVSPPPQAPAFVARTAS 309
Dh 372 RQRPPLPPQPPPVNLAS 390

RESULT 2
075160 PRELIMINARY; PRT; 818 AA.
AC 075160:
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIA0672.
GN KIA0672.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohata O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014572; BAA31647.1; -.
DR HSSP; 007960; IRGP.
DR InterPro; IPR000198; RhogAP.
DR Pfam; PF00620; RhogAP.1.
DR SMART; SM00324; RhogAP.1.
SQ SEQUENCE 818 AA; 89262 MW; 996C8ECF9B523808 CRC64;

Query Match 7.3%; Score 126; DB 4; Length 818;
Best Local Similarity 20.7%; Pred. No. 0.012;
Matches 80; Conservative 55; Mismatches 140; Indels 112; Gaps 16;

Qy 3 LPPFSCVYGVGVYVVP--DGIVNLSKSCSVH-----DLLGNPDYVCGSPPTTYTN S5
Dh 397 MPPSMALVLD-PULLPQAEKNTBMTYSLQIVILPIIQAHDWFFGCELEFNITG 455
Qy 456 RMAAEVLEVGVYVVCATPPNAQPIELERQPKVYVHR-----GSKILPFSKHGVHVELR 109
Dh 456 NYGSPVYVHNHNNYSSMSPSPMDPADRRQPEQARRLSVATDNMKLEFYKDKG--LRKIQ 513
Qy 110 S-----PTNGSQOSKRFEDHSVTYMQOLEISINEGPEPHLAGDPSKHLKLVFI 158
Dh 514 SNQGVNMTNNVARRGSSAGKRV-----SCAPSPMQPPAPPAELAAPLPSF----- 559
Qy 159 RHCRLALPLPTISIDLMESFLNISEALSP-----TATAKDEITQMLISAAR----- 207
Dh 560 -----LPEQPLDSPAAPALPSVSLGLQPEPPEPTSTTKSK--ELSPGSAQKGSBSSQ 609
```

```
Qy 208 -----SELGMYVKRQEFYLRARRRRKFPAMKPVLOSISEMKPYMEFHTPM 253
Dh 610 GTACAGTQPGAGPQAGAPGASPSGSPADQSPHTLRK-----VSKKLAIPIPKVVFQGP 664
Qy 254 AYRDSG---SPPKNAST-PSLGPKNISPPROVSVP-----Q 286
Dh 665 AAMDQSAQGLSPVSLSPPTPEPTSPICGLSPGCGYSLASGQLSPAAPPLASPSVFTSTLS 724
Qy 287 RSSPPPK---NVSPPPQAPAFVARTAS 309
Dh 725 KSRPTPKPRQRPPLPPQPPPVNLAS 751

RESULT 3
09VZT8 PRELIMINARY; PRT; 1381 AA.
AC 09VZT8:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG14964.
GN CG14964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRATNE-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Wortman J.R., Riordan M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernasconi B.P., Bhandari D., Bolshakov S.,
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekurov G., Milshina N.V., Moberly C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-T., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zhi H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2198(2000).
DR EMBL; AE003477; AAF47729.1; -.
DR HSSP; PS6276; 1TUK.
DR FlyBase; Fggn0035410; CG14964.
```


RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 402:761-768(1999).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC006585; AAD3008.1; -
 DR InterPro: IPR002965; P_rich_extensn.
 DR PRINTS: PRO1217; PRICHEXTENSN.
 SQ SEQUENCE 742 AA; 81758 MW; EB5E7A4096F57D28 CRC64;

Query Match 6.4%; Score 114; DB 10; Length 742;
 Best Local Similarity 21.7%; Pred. No. 0.12;
 Matches 60; Conservative 48; Mismatches 127; Indels 42; Gaps 9;

OY 52 TTNRMAREVLEVGTVFYCAAPPNAPLERQPKYVHRSKILPFSKHGVHRELKSP 111
 Db 352 TVENQJTAANILOESLDGSPSPDSRLLESALKVEKTKLM-----ISENIVSSP 404
 OY 112 THGSGOSRRKVPDHSVTMQOLESTRNEGPEPHLADRSKHLKVFIRHCLALRLPRIS 171
 Db 405 DTSSPKER-----DTSSHKSTAD--PNSILTKVDESRLGRLVSVQRVHSHKISPPMV 456
 OY 172 IDIMSEPLPNLSGELSPFATANDETITOMILKSAARSELGM-----YVSKQEFYLRRA 225
 Db 457 QSPVTSPLPRLPTGSPASISRFH-----SSPSLIGITITLHDHSGCDESTSS 508
 OY 226 RRRRRKPAKPVLOSISEMKPMEF-----HTPMAYDSQSPKNAASPISLGRKNTSPRQ 281
 Db 509 PASPSISFLPTLHLPTSSQPKKASPCQSPSPPVHNSNGPFAEAAYTSPLP-LKPLRI 567
 OY 282 VSVPOSSSPKPNVSPPOPAFAFARTASKYSAASQV 318
 Db 568 LSRP-----PP-----PPPPISLKRSTPSPSTNSI 596

RESULT 9
 Q9T240 PRELIMINARY; PRT; 539 AA.
 ID Q9T240;
 AC Q9T240;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOPHETICAL 62.1 KDA PROTEIN.
 GN W0783.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 283:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Wohldmann P., Bauer C., Rohlfing T., Gilliam B.;
 RT "The sequence of *C. elegans* cosmid W0783.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;

RT "Direct Submission";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF100304; AAC68911.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 539 AA; 62094 MW; F4A47C536DCF179B CRC64;

Query Match 6.4%; Score 111.5; DB 5; Length 539;
 Best Local Similarity 23.0%; Pred. No. 0.13;
 Matches 43; Conservative 38; Mismatches 75; Indels 31; Gaps 6;

OY 154 KLVFIRHCLRLRLPRISID-IMESPLPNLSGELSPFATANDETITOMILKSAARSELGM 212
 Db 300 KRYVNRRTTEAFRRRRMEEMLLAMPGRINGEAL-----DRAQIVYNNARAEAR 352
 OY 213 YVSKQEFYLRARRRRKPAKPVLOSISEMKPVMEFHTPMAYRDSQSPPKNASTPLPG 272
 Db 353 LRYQMTPTDQRKSYNKRYPKRRREMDMISAG-----MAYGSGITGTGTPAKKD 408
 OY 273 P-----KNISPPROVSVPOR---SSPPKAVSPPOPAFAFARTASKYSAASQ 317
 Db 409 DMDALSLERVYKRTQHQAOJILNRANOSSGTPKSVTPPOAVTISNGQOH---QQ 464
 OY 318 VQRNRGN 324
 Db 465 LVGNQGN 471

RESULT 10
 O04712 PRELIMINARY; PRT; 202 AA.
 ID O04712;
 AC O04712;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE T10M13.10.
 GN T10M13.10 OR AGAA.5 OR AT4G02090.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Johnson A.F., la Bastide M., Lodhi M., Hoffman J., Hasegawa A.,
 RA gnoj L., Gottesman T., Granat S., Hameed A., Kaplan N., Schutz K.,
 RA Shohdy N., Van Keuren K., Parnell L., Dedhia N., Martienssen R.,
 RA McCombie W.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. LANDSBERG;
 RA Till S., Granat S., Parnell L., Kaplan N., Hoffman J., Lodhi M.,
 RA Johnson A.F., Dedhia N., Martienssen R., McCombie W.R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF001308; AAC78701.1; -
 DR EMBL: AF001535; AAB57800.1; -
 DR EMBL: AL161493; CAB80702.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 202 AA; 22316 MW; CAA7BEDF62FAA8AF CRC64;

Query Match	6.48;	Score 111;	DB 10;	Length 202;
Best Local Similarity	21.98;	Pred. No. 0.041;		
Matches 54;	Conservative 24;	Mismatches 89;	Indels 80;	Gaps 7;

```

QY      16 RVPVDPDASHVNLKSSCSVHDDLGNPDYVCGSTP-----YITNRMA---AAEVL 65
      19 KVLISDERVONLEEEFTVAELMLELNPQHVHVEEDSSISFNDAKIVKPKRLPLPAKRTL 78
QY      64 EYGVLYEVCATPNAQPFLEERQPKVHHGSKILPRFSGHGVHVELRSPHGSQGSRYVFD 123
      79 EPGKTYLV-----LPA-KRGGRAAKSSSAVLITSEEMKRL- 113
Db      124 YHSVYMOALESIRNEGEPHLADRPKHLKLVFRHCLRLRLPRISIDLMESEPLNLS 183
      114 -----FSAITAMVHSSPYEGEILMPFT 135
QY      184 GSAISPTATKADDETOMILKSAARSELGMVYNSKROEFLTPARRRRKFMKRVLOSISEM 243
      136 TMSYKNNNPATD--TVAAVTSYGRLEAMEEEDREFEFLSKQLSR---GMRKSLDPIREK 190
Db
QY      244 KRVMEFH 250
      191 KAKKKIH 197

```

RESULT	11			
091T14				
ID	091T14	PRELIMINARY:	PRT:	393 AA.
AC	091T14:			
DT	01-DEC-2001 (TEMBLrel. 19, Created)			
DT	01-DEC-2001 (TEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)			
DE	117,			
OS	Tupala herpesvirus.			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
CC	Betaherpesvirinae.			
OX	NCBI_TaxID:10397;			
RN	11]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-2:			
RX	MEDLINE:21211637; PubMed-11312357;			
RA	Bahr U., Darl G.;			
RT	Analysis and Characterization of the Complete Genome of Tupala (Tree			
RT	Shrew) Herpesvirus."			
RL	J. Virol. 75:4854-4870(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-2:			
RC	Darl G., Bahr U.;			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: AF281817; ANK57163.1. -			
SO	SEQUENCE 393 AA; 41784 MW; F48A27BB3C53BDF CRC64;			

Query Match	6.38;	Score 109.5;	DB 12;	Length 393;
Best Local Similarity	22.78;	Pred. No. 0.13;		
Matches	45;	Conservative	25;	Mismatches 53;
				Indels 75;
				Gaps 9;

```

QY 140 PEHLHAGRGPEKH-----LKIVFIR-----HCL-RALRIPRISIDLESELP 180
Db 68 PPPPPDGGDPRHRGTPPTVTITTYTVTVRAPASPTHAVERTRNSP-----ESPP 120

QY 181 NISGE-----ALSPRTAKDEITOMILKSANSEJGMYVSKROEFLYLRARRR 225
Db 121 SPFAEPTAIVASAPRLGPALEVEGSGIHPRRLRSTPRAELGALLE----- 165

QY 230 KFAKMPVLQSISEKKPMVEFHHPMAVYRDSGSPKRNASTSLPGKNISPPROYSPQSS 285
Db 168 --VLKPIEASVSGI-----QTAPDTPPTPR--SPPR--SPPKRAI 202

QY 290 PPPKNVSPPPQAFVAVT 307

```

Db 203 PPPPPPPPPPAET 220

RESULT	12	
09FHXL		
ID	PRELIMINARY;	PRT; 1073 AA.
OC	09FHXL	
DT	01-MAR-2001 (TREMblrel. 16, Created)	
DT	01-MAR-2001 (TREMblrel. 16, Last sequence update)	
DT	01-DEC-2001 (TREMblrel. 19, Last annotation update)	
DE	TMV RESISTANCE PROTEIN-LIKE.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
CC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;	
CC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
NP	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=COLUMBIA;	
RX	MEDLINE=99397451; PubMed=10470850;	
RA	Kaneo T., Katon T., Sato S., Nakamura Y., Asamizu E., Kotani H.,	
RA	Miyajima N., Tabata S.;	
RT	Structural analysis of Arabidopsis thaliana chromosome 5. IX.	
RT	Sequence features of the regions of 1,011,550 bp covered by seventeen	
RL	PL and TAC clones.";	
DR	PL and TAC clones.";	
DR	EMBL: AB017067; BAB06447.1; -	
DR	InterPro: IPR0000345; CysC_heme_bind.	
DR	InterPro: IPR000408; CysC_heme_bind.	
DR	InterPro: IPR000306; ZnF_FYVE.	
DR	Pfam: PF01363; FYVE; 1.	
DR	Pfam: PF00412; RCL1; 6.	
DR	PRINTS: PR00633; RCLNDNSATTON.	
DR	SMART: SM00064; FYVE; 1.	
DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.	
DR	PROSITE: PS00012; RCL13; 6.	
QO	SEQUENCE 1073 AA; 116497 MW; 7.96FC2P5D4A0CD4 CRC64;	

Query Match	6.38;	Score 109.5;	DB 10;	Length 1073;
Best Local Similarity	23.68;	Pred. No. 0.49;		
Matches	89;	Conservative	46;	Mismatches 147;
				Indels 95;
				Gaps 16;

```

OY      1 MGITPFCSCV-----TWQGVVRVVRBDHV-----EUKSCSSVHDLULNIPYVYCSIF 30
      : : : : :
Db      507 VGLITTSGVYVMGSTVYGQGLGNPNADKELICLVEKLTIKDC-VEELACAGVHAVALTSRN 555
      : : : : :
OY      51 YTTIRMAAEVLEYCYVTCVFCATPNQPLEKOPVVHRSKILTFPSKHGVHVELRS 110
      : : : : :
Db      566 EYVTPGKCANRRGHGVHEDVEDKAPTLVDALKERHYNVNIACGSNFTALCIH-----KWS 620
      : : : : :
OY      111 PRHSGOS--RKVEDY-----HSYMOOLEISr-----NEGEPHLGDRPS 150
      : : : : :
Db      621 GTEDGQSCACQAGFTRRKRNHCNCGVLVHCSCSKSLAKAALPNPG-KYRYCD--S 677
      : : : : :
OY      151 KHLKLVFIRHOLRLALPRTSIDLMESTPISLGE-----ALSPATKDEIT 198
      : : : : :
Db      678 CHSKL-----SKVSEANIDSKNNVNRLEGEENKDRDKTETIRLANSGLPSMIDLIK 728
      : : : : :
OY      199 QMILSAARSLGLVYVSKQDEYLRARARRRRKAMKPYVLOSISMKVMYMFITPRAYHDS 258
      : : : : :
Db      729 QLDNRRAQGR-----KADTSLVRSQPLTQLDKDALTNVALDR----- 768
      : : : : :
OY      259 GSPFKNASTPRLPEKPNISPPQVSVPPRSSPPPKVNSPPQ-----PAFARASTKS 312
      : : : : :
Db      769 RGCPKPAATP-----SSSPASVPSPRSNSPPRSVYETPLANVLGQFSTJAEBSLAKITN 820

```

RESULT 13

09UHA8
ID 09UHA8 PRELIMINARY: PRT: 2296 AA.
AC 09UHA8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SPLICING COACTIVATOR SUBUNIT SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20132238; PubMed=10668604;
RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosonina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits.";
RL RNA 6:111-120(2000).
DR EMBL: AF201422, AAF21439.1;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
DR SEQUENCE 2296 AA; 251964 MW; 17C0BD4EAL0A9CF9 CRC64;

Query Match 6.3%; Score 109.5; DB 4; Length 2296;
Best Local Similarity 26.5%; Pred. No. 1.3;
Matches 53; Conservative 27; Mismatches 75; Indels 45; Gaps 10;

OY 162 LRRLRLPRISIDIMESPLNLSGALSPATAKDEI---TOMIL--KSAARSELGMVYSK 216
DB 279 LEVLOAKTLQLEWIGVPLDQDAREMLSVNVLPTQRASSPEATKPPSPYDCK 338
OY 217 ROEYLRRARRRRKFAKMP-VLOSISEMKPVMEFTTP-MAYRDSGSPKNASTPSLPPK 274
DB 339 DKP-----KKEKATPSPSPERSSTGPEPPAPTPLLARHGSGQPLATTPIISQEP- 390
OY 275 NISPPROVSVPORSSPPPKN-----VSPR-PQPAFVARTAS----- 309
DB 391 -VAPPEAS-PRDRSPFKSPKLPDSSSESSPPSPQPKVSHHASSSPSPKAPAPG 448
OY 310 ---KYSASQOVORNGNAK 326
DB 449 SHREISSPTSKNRHGRK 468

RESULT 14
O9M2X4
ID O9M2X4 PRELIMINARY: PRT: 651 AA.
AC O9M2X4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 72.3 KDA PROTEIN.
GN T16K5.190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA Hewes H.W., Lemcke K., Mayer K.F.X., Quetler F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132965; CAB66922.1;
KW Hypothetical protein.
SQ SEQUENCE 651 AA; 72344 MW; F4ED1138046F2F1D CRC64;

Query Match 6.3%; Score 109; DB 10; Length 651;
Best Local Similarity 22.5%; Pred. No. 0.28;
Matches 62; Conservative 30; Mismatches 92; Indels 92; Gaps 13;

OY 52 TITNRMAAEVLEYGVTVFCATPNAOPFLEROKPVHRS--KILPRSKHGVEELR 109
DB 387 SISRKILQRIYEDDETESCS-----PGLIHATSFASPLPRGSKKILRTVPLS 435
OY 110 -----SPTHSQOSKRVFVYSVTMOOLESI RNEGEPHLADNRSHKLVITRCLR 163
DB 436 ESDIRPSSFHKTPEGRR-----KLTKELQKFTKESTE-----N 469
OY 164 ALRLRLPRISIDIMESPLNLSGALSPATAKDEITOMILKSAAR--SELGMVYSKROEF 220
DB 470 ALK-----ELVSSPEEGSGSE-----KQAAKGTSEPVGECKKKQ- 506
OY 221 YLRARRRRKFAKMPVLOSISEMKPVMEFTTPM----AYRDSGSPKNASTPS-LPGPKN 275
DB 507 -----RVRYKRRNKPSDKSITEKKKMSYODQHPVSAPPQGYPPKGGYPPAG 561
OY 276 ISPPR--QVSVPORSSPPK-----NVSPPQ 300
DB 562 YPPPOYPQAGYPPAGYPPPOGCGGYPAQGYPPQ 597

RESULT 15
Q9EEF8
ID Q9EEF8 PRELIMINARY: PRT: 517 AA.
AC Q9EEF8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL P60 PROTEIN.
OS Anticarsia gemmatilis nuclear polyhedrosis virus (AgmNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=31507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2D;
RX MEDLINE=92268862; PubMed=1588315;
RA Zanotto P.M., Sampaio M.J., Johnson D.W., Rocha T.L., Marunjak J.E.;
RT "The Anticarsia gemmatilis nuclear polyhedrosis virus polyhedrin gene
region: sequence analysis, gene product and structural comparisons.";
RL J. gen. Virol. 73:1049-1056(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-2D;
RX Zanotto P.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-2D;
RX MEDLINE=94044871; PubMed=8228320;
RA Zanotto P.M.A., Kessing B.D., Marunjak J.E.;
RT "Phylogenetic interrelationships among baculoviruses: Evolutionary
rates and host associations.";
RL J. Invertebr. Pathol. 62:147-164(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-2D;
RX MEDLINE=20033815; PubMed=10564750;
RA Chen X., Ijzkel W.F.J., Dominy C., Zanotto P.M.A., Hashimoto Y.,
RA Fakoor O., Hayakawa T., Wang C.H., Premunkar A., Krell P.D., Hu Z.,
RA Viak J.M.;
RT "Identification, sequence analysis and phylogeny of the Ief-2 gene of
RT Helicoverpa armigera single-nucleocapsid baculovirus.";
RL Virus Res. 65:21-32(1999).
DR EMBL: Y17753; CAC03567.1;
DR InterPro: IPR003124; WH2.
DR Pfam: PF02205; WH2; 2.
DR SMART: SM00246; WH2; 2.

SEQ SEQUENCE 517 AA: 59000 MW: B50872BBF3633995 CRC64;

Query Match 6.2%: Score 108; DB 12; Length 517;

Best Local Similarity 20.9%; Pred. No. 0.26;

Matches 73; Conservative 40; Mismatches 102; Indels 134; Gaps 15;

QY 24 VENLSKGSVHD--LLGMPDYVCGSPYITTRMAAEVLECYTFY-----C 72
DB 110 VENILRSDAVDVETIVLSGDFE-----EYPNTWVNTFERTRCRC 151
QY 73 ATPNAOPFLERQPKV---VHRGSKI-LPFSKHGVHRELSPTHGSGQSRKVFYDSY 127
DB 152 RPQQLRLYLKQRRKLMCLHLHPLKRLQYLHPSLH-KHLHPLNRLQH-----FHPL 200
QY 128 TMOQLFSTRNEGPEPLADGRPSKHLKVFIRHCIRALRLPRISIDLMESPLPULSGEAL 187
DB 201 NRLQ-----HLHLNQLQHLHLSPKRHHRLKRLQRL-----L 235
QY 188 SPTATANDEITQMLKSAARSELGMYVSKROEFLRRARRRRKFAWKPVLOSISEMKPYM 247
DB 236 YP-----LKH-----FYLSPHQ-HLHPLRHLHRLHMLHPLDLLQNRNPRV 274
QY 248 EFH-----TPMAYRDSGSPPKNAST 267
DB 275 RFQLOKVRCLARLTNLNLTLELALMDKPTSVPPKPPHLKSNAMFVGDKIAPFCTNT 334
QY 268 PSLPGPKNISPPROVSYPQSSPPPKNYSPPQAFAYAR-TASKXSAS 315
DB 335 P--PPNATTSFOVGSMPPPVAPPNNVMPASASAAKHAANKHAAAT 381

Search completed: September 16, 2002, 00:51:11
Job time: 475 sec



Blast 2 Sequences results

PubMed

Entrez

BLAST

OMIM

Taxonomy

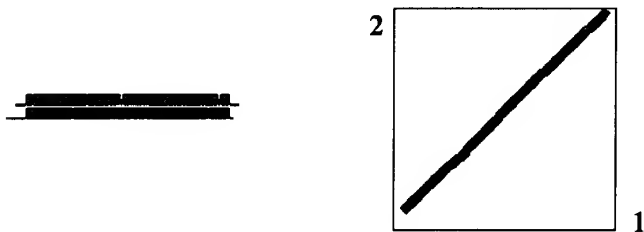
Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.3 [Apr-24-2002]

Matrix BLOSUM62 gap open: 11 gap extension: 1
 x_dropoff: 50 expect: 300. wordsize: 3

Sequence 1 gi 2851522 Mitogen-activated protein kinase 12 (Extracellular
 signal-regulated kinase 6) (ERK-6) (ERK5)
 (Stress-activated protein kinase-3) (Mitogen-activated
 protein kinase p38 gamma) (MAP kinase p38 gamma). **Length** 367 (1 .. 367)

Sequence 2 gi 1136298 WIPK [Nicotiana tabacum] **Length** 375 (1 .. 375)



NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

Score = 292 bits (747), Expect = 5e-78
 Identities = 149/339 (43%), Positives = 219/339 (63%), Gaps = 8/339 (2%)

```

Query: 21 WEVRAVYRD-LQPVGSGAYGAVCSAVDGRGAKVAIKKLYRPFQSEFAKRAYRELRLLK 79
      +E+  YR  + P+G GAYG VCS ++          VA+KK+  F   + AKR  RE++LL+
Sbjct: 36 FEITTKYRPPIMPIGRGAYGIVCSVLNTELNEMVAVKKIANAFDIYMDAKRTLREIKLLR 95

Query: 80 HMRHENVIGLLDVFTPDETLDDFTDFYLVMPFMGTDLGKLMK-HEKLGEDRIQFLVYQML 138
      H+  HENVIGL DV  P   +F+D Y+   M TDL ++++ ++ L ED  Q+ +YQ+L
Sbjct: 96 HLDHENVIGLRDVIPPP-LRREFSDVYIATELMDTDLHQIIRSNQGLSEDHCQYFMYQLL 154

Query: 139 KGLRYIHAAGIIHRDLKPGNLAVNEDCELKILDFGLARQ--ADSEMTGYVVTRWYRAPEV 196
      +GL+YIH+A ++HRDLKP NL VN +C+LKI DFGLAR   + MT YVVTRWYRAPE+
Sbjct: 155 RGLKYIHSANVLHRDLKPSNLLVNANCDLKICDFGLARPNIEENMTEYVVTRWYRAPEL 214

Query: 197 ILNWMRYTQTVDIWSVGCIMAEMITGKTLFGSDHLDQLKEIMKVGTGPPAEFVQRLQSD 256
      +LN  YT  +D+WSVGC I E++  K LF G DH+ Q++ + ++ GTP   + LQ++
Sbjct: 215 LLNSSDYTAIDVWSVGCIFMELMNRKPLFGGKDHVHQIRLLTELLGTPTEADLGFLQNE 274

Query: 257 EAKNYMKGLEPELEKKDFASILTNASPLAVNLEKMLVLDAEQRTAGEALAHYPYFESLHD 316
      +AK Y++ LP+ ++ A + + +PLA++L++KML D +R+T EAL HPY  LHD
Sbjct: 275 DAKRYIRQLPQHPRQQLAEVFPHVNPLAIDLVDKMLTFDPTRRITVEEALDHPYLAKLHD 334

Query: 317 TEDEPQVQKYDDSFDDVDRTL--DEWKRVTYKEVLSFKP 353
      DEP +   SFD  + + ++ K + Y+E LS  P
Sbjct: 335 AGDEP-ICVPVFSDFEQGIGEEQIKDMIYQEALSLNP 372

CPU time:      0.08 user secs.      0.01 sys. secs      0.09 total secs.
  
```


Lambda	K	H
0.320	0.137	0.405

Gapped Lambda	K	H
0.267	0.0410	0.140

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Number of Hits to DB: 1022

Number of Sequences: 0

Number of extensions: 86

Number of successful extensions: 4

Number of sequences better than 300.0: 1

Number of HSP's better than 300.0 without gapping: 1

Number of HSP's successfully gapped in prelim test: 0

Number of HSP's that attempted gapping in prelim test: 0

Number of HSP's gapped (non-prelim): 1

length of query: 367

length of database: 330,426,180

effective HSP length: 123

effective length of query: 244

effective length of database: 219,683,991

effective search space: 53602893804

effective search space used: 53602893804

T: 9

A: 40

X1: 16 (7.4 bits)

X2: 129 (49.7 bits)

X3: 129 (49.7 bits)

S1: 41 (21.8 bits)

S2: 60 (27.7 bits)

